

**METHODS OF DETECTION AND NOTIFICATION OF BIOAGENT  
CONTAMINATION**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

5           This application claims priority to U.S. provisional application Serial No. 60/453,607 filed March 10, 2003 and to U.S. provisional application Serial No. 60/454,165 filed March 12, 2003, each which is incorporated herein by reference in its entirety.

**REFERENCE TO GOVERNMENT GRANTS**

10           The present invention was supported by funds from the U.S. Government (DARPA/SPO contract No. BAA00-09) and the U.S. Government may therefore have certain rights in the invention.

**FIELD OF THE INVENTION**

15           The present invention relates generally to the field of investigational bioinformatics and more particularly to methods of detection and notification of bioagent contamination. Detection and notification of bioagent contamination is important for determining a proper course of treatment, irradiation, and/or containment of the bioagent in such cases as biological warfare.

20

**BACKGROUND OF THE INVENTION**

          In the United States, hospitals report well over 5 million cases of recognized infectious disease-related illnesses annually. Significantly greater numbers remain undetected, both in the inpatient and community setting, resulting in substantial morbidity and mortality. Critical intervention for infectious disease relies on rapid, sensitive and specific detection of the offending pathogen, and is central to the mission of microbiology laboratories at medical centers. Unfortunately, despite the recognition that outcomes from infectious illnesses are directly associated with time to pathogen recognition, as well as accurate identification of the class and species of microbe, and ability to identify the presence of drug resistance isolates, conventional hospital laboratories often remain encumbered by traditional slow multi-step culture based assays. Other limitations of the conventional laboratory which have become increasingly apparent include: extremely prolonged wait-times for pathogens with long generation time (up to several weeks); requirements for

additional testing and wait times for speciation and identification of antimicrobial resistance; diminished test sensitivity for patients who have received antibiotics; and absolute inability to culture certain pathogens in disease states associated with microbial infection.

For more than a decade, molecular testing has been heralded as the diagnostic tool for the new millennium, whose ultimate potential could include forced obsolescence of traditional hospital laboratories. However, despite the fact that significant advances in clinical application of PCR techniques have occurred, the practicing physician still relies principally on standard techniques. A brief discussion of several existing applications of PCR in the hospital-based setting follows.

Generally speaking molecular diagnostics have been championed for identifying organisms that cannot be grown *in vitro*, or in instances where existing culture techniques are insensitive and/or require prolonged incubation times. PCR-based diagnostics have been successfully developed for a wide variety of microbes. Application to the clinical arena has met with variable success, with only a few assays achieving acceptance and utility.

One of the earliest, and perhaps most widely recognized applications of PCR for clinical practice is in detection of *Mycobacterium tuberculosis*. Clinical characteristics favoring development of a nonculture-based test for tuberculosis include week to month long delays associated with standard testing, occurrence of drug-resistant isolates and public health imperatives associated with recognition, isolation and treatment. Although frequently used as a diagnostic adjunctive, practical and routine clinical application of PCR remains problematic due to significant inter-laboratory variation in sensitivity, and inadequate specificity for use in low prevalence populations, requiring further development at the technical level. Recent advances in the laboratory suggest that identification of drug resistant isolates by amplification of mutations associated with specific antibiotic resistance (e.g., *rpoB* gene in rifampin resistant strains) may be forthcoming for clinical use, although widespread application will require extensive clinical validation.

One diagnostic assay, which has gained widespread acceptance, is for *C. trachomatis*. Conventional detection systems are limiting due to inadequate sensitivity and specificity (direct immunofluorescence or enzyme immunoassay) or the requirement for specialized culture facilities, due to the fastidious characteristics of this microbe. Laboratory development, followed by widespread clinical validation testing in a variety of acute and nonacute care settings have demonstrated excellent sensitivity (90-100%) and specificity (97%) of the PCR assay leading to its commercial development. Proven efficacy of the PCR

assay from both genital and urine sampling, have resulted in its application to a variety of clinical setting, most recently including routine screening of patients considered at risk.

While the full potential for PCR diagnostics to provide rapid and critical information to physicians faced with difficult clinical-decisions has yet to be realized, one recently developed assay provides an example of the promise of this evolving technology. Distinguishing life-threatening causes of fever from more benign causes in children is a fundamental clinical dilemma faced by clinicians, particularly when infections of the central nervous system are being considered. Bacterial causes of meningitis can be highly aggressive, but generally cannot be differentiated on a clinical basis from aseptic meningitis, which is a relatively benign condition that can be managed on an outpatient basis. Existing blood culture methods often take several days to turn positive, and are often confounded by poor sensitivity or false-negative findings in patients receiving empiric antimicrobials. Testing and application of a PCR assay for enteroviral meningitis has been found to be highly sensitive. With reporting of results within 1 day, preliminary clinical trials have shown significant reductions in hospital costs, due to decreased duration of hospital stays and reduction in antibiotic therapy. Other viral PCR assays, now routinely available include those for herpes simplex virus, cytomegalovirus, hepatitis and HIV. Each has a demonstrated cost savings role in clinical practice, including detection of otherwise difficult to diagnose infections and newly realized capacity to monitor progression of disease and response to therapy, vital in the management of chronic infectious diseases.

The concept of a universal detection system has been forwarded for identification of bacterial pathogens, and speaks most directly to the possible clinical implications of a broad-based screening tool for clinical use. Exploiting the existence of highly conserved regions of DNA common to all bacterial species in a PCR assay would empower physicians to rapidly identify the presence of bacteremia, which would profoundly impact patient care. Previous empiric decision making could be abandoned in favor of educated practice, allowing appropriate and expeditious decision-making regarding need for antibiotic therapy and hospitalization.

Experimental work using the conserved features of the 16S rRNA common to almost all bacterial species, is an area of active investigation. Hospital test sites have focused on "high yield" clinical settings where expeditious identification of the presence of systemic bacterial infection has immediate high morbidity and mortality consequences. Notable clinical infections have included evaluation of febrile infants at risk for sepsis, detection of

bacteremia in febrile neutropenic cancer patients, and examination of critically ill patients in the intensive care unit. While several of these studies have reported promising results (with sensitivity and specificity well over 90%), significant technical difficulties (described below) remain, and have prevented general acceptance of this assay in clinics and hospitals (which remain dependent on standard blood culture methodologies). Even the revolutionary advances of real-time PCR technique, which offers a quantitative more reproducible and technically simpler system, remains encumbered by inherent technical limitations of the PCR assay.

The principle shortcomings of applying PCR assays to the clinical setting include: inability to eliminate background DNA contamination; interference with the PCR amplification by substrates present in the reaction; and limited capacity to provide rapid reliable speciation, antibiotic resistance and subtype identification. Some laboratories have recently made progress in identifying and removing inhibitors; however background contamination remains problematic, and methods directed towards eliminating exogenous sources of DNA report significant diminution in assay sensitivity. Finally, while product identification and detailed characterization has been achieved using sequencing techniques, these approaches are laborious and time-intensive thus detracting from its clinical applicability.

Rapid and definitive microbial identification is desirable for a variety of industrial, medical, environmental, quality, and research reasons. Traditionally, the microbiology laboratory has functioned to identify the etiologic agents of infectious diseases through direct examination and culture of specimens. Since the mid-1980s, researchers have repeatedly demonstrated the practical utility of molecular biology techniques, many of which form the basis of clinical diagnostic assays. Some of these techniques include nucleic acid hybridization analysis, restriction enzyme analysis, genetic sequence analysis, and separation and purification of nucleic acids (See, e.g., J. Sambrook, E. F. Fritsch, and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989). These procedures, in general, are time-consuming and tedious. Another option is the polymerase chain reaction (PCR) or other amplification procedure that amplifies a specific target DNA sequence based on the flanking primers used. Finally, detection and data analysis convert the hybridization event into an analytical result.

Other not yet fully realized applications of PCR for clinical medicine is the identification of infectious causes of disease previously described as idiopathic (e.g.

*Bartonella henselae* in bacillary angiomatosis, and *Tropheryma whippellii* as the uncultured bacillus associated with Whipple's disease). Further, recent epidemiological studies which suggest a strong association between Chlamydia pneumonia and coronary artery disease, serve as example of the possible widespread, yet undiscovered links between pathogen and host which may ultimately allow for new insights into pathogenesis and novel life sustaining or saving therapeutics.

For the practicing clinician, PCR technology offers a yet unrealized potential for diagnostic omnipotence in the arena of infectious disease. A universal reliable infectious disease detection system would certainly become a fundamental tool in the evolving diagnostic armamentarium of the 21<sup>st</sup> century clinician. For front line emergency physicians, or physicians working in disaster settings, a quick universal detection system, would allow for molecular triage and early aggressive targeted therapy. Preliminary clinical studies using species specific probes suggest that implementing rapid testing in acute care setting is feasible. Resources could thus be appropriately applied, and patients with suspected infections could rapidly be risk stratified to the different treatment settings, depending on the pathogen and virulence. Furthermore, links with data management systems, locally regionally and nationally, would allow for effective epidemiological surveillance, with obvious benefits for antibiotic selection and control of disease outbreaks.

For the hospitalists, the ability to speciate and subtype would allow for more precise decision-making regarding antimicrobial agents. Patients who are colonized with highly contagious pathogens could be appropriately isolated on entry into the medical setting without delay. Targeted therapy will diminish development of antibiotic resistance. Furthermore, identification of the genetic basis of antibiotic resistant strains would permit precise pharmacologic intervention. Both physician and patient would benefit with less need for repetitive testing and elimination of wait times for test results.

It is certain that the individual patient will benefit directly from this approach. Patients with unrecognized or difficult to diagnose infections would be identified and treated promptly. There will be reduced need for prolonged inpatient stays, with resultant decreases in iatrogenic events.

Mass spectrometry provides detailed information about the molecules being analyzed, including high mass accuracy. It is also a process that can be easily automated. Low-resolution MS may be unreliable when used to detect some known agents, if their spectral lines are sufficiently weak or sufficiently close to those from other living organisms

in the sample. DNA chips with specific probes can only determine the presence or absence of specifically anticipated organisms. Because there are hundreds of thousands of species of benign bacteria, some very similar in sequence to threat organisms, even arrays with 10,000 probes lack the breadth needed to detect a particular organism.

5           Antibodies face more severe diversity limitations than arrays. If antibodies are designed against highly conserved targets to increase diversity, the false alarm problem will dominate, again because threat organisms are very similar to benign ones. Antibodies are only capable of detecting known agents in relatively uncluttered environments.

10           Several groups have described detection of PCR products using high resolution electrospray ionization-Fourier transform-ion cyclotron resonance mass spectrometry (ESI-FT-ICR MS). Accurate measurement of exact mass combined with knowledge of the number of at least one nucleotide allowed calculation of the total base composition for PCR duplex products of approximately 100 base pairs. (Aaserud *et al.*, *J. Am. Soc. Mass Spec.*, **1996**, *7*, 1266-1269; Muddiman *et al.*, *Anal. Chem.*, **1997**, *69*, 1543-1549; Wunschel *et al.*, *Anal.*  
15 *Chem.*, **1998**, *70*, 1203-1207; Muddiman *et al.*, *Rev. Anal. Chem.*, **1998**, *17*, 1-68). Electrospray ionization-Fourier transform-ion cyclotron resistance (ESI-FT-ICR) MS may be used to determine the mass of double-stranded, 500 base-pair PCR products via the average molecular mass (Hurst *et al.*, *Rapid Commun. Mass Spec.* **1996**, *10*, 377-382). The use of matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry  
20 for characterization of PCR products has been described. (Muddiman *et al.*, *Rapid Commun. Mass Spec.*, **1999**, *13*, 1201-1204). However, the degradation of DNAs over about 75 nucleotides observed with MALDI limited the utility of this method.

          U.S. Patent No. 5,849,492 describes a method for retrieval of phylogenetically informative DNA sequences which comprise searching for a highly divergent segment of  
25 genomic DNA surrounded by two highly conserved segments, designing the universal primers for PCR amplification of the highly divergent region, amplifying the genomic DNA by PCR technique using universal primers, and then sequencing the gene to determine the identity of the organism.

          U.S. Patent No. 5,965,363 discloses methods for screening nucleic acids for  
30 polymorphisms by analyzing amplified target nucleic acids using mass spectrometric techniques and to procedures for improving mass resolution and mass accuracy of these methods.

WO 99/14375 describes methods, PCR primers and kits for use in analyzing preselected DNA tandem nucleotide repeat alleles by mass spectrometry.

WO 98/12355 discloses methods of determining the mass of a target nucleic acid by mass spectrometric analysis, by cleaving the target nucleic acid to reduce its length, making the target single-stranded and using MS to determine the mass of the single-stranded shortened target. Also disclosed are methods of preparing a double-stranded target nucleic acid for MS analysis comprising amplification of the target nucleic acid, binding one of the strands to a solid support, releasing the second strand and then releasing the first strand which is then analyzed by MS. Kits for target nucleic acid preparation are also provided.

PCT WO97/33000 discloses methods for detecting mutations in a target nucleic acid by nonrandomly fragmenting the target into a set of single-stranded nonrandom length fragments and determining their masses by MS.

U.S. Patent No. 5,605,798 describes a fast and highly accurate mass spectrometer-based process for detecting the presence of a particular nucleic acid in a biological sample for diagnostic purposes.

WO 98/21066 describes processes for determining the sequence of a particular target nucleic acid by mass spectrometry. Processes for detecting a target nucleic acid present in a biological sample by PCR amplification and mass spectrometry detection are disclosed, as are methods for detecting a target nucleic acid in a sample by amplifying the target with primers that contain restriction sites and tags, extending and cleaving the amplified nucleic acid, and detecting the presence of extended product, wherein the presence of a DNA fragment of a mass different from wild-type is indicative of a mutation. Methods of sequencing a nucleic acid via mass spectrometry methods are also described.

WO 97/37041, WO 99/31278 and U.S. Patent No. 5,547,835 describe methods of sequencing nucleic acids using mass spectrometry. U.S. Patent Nos. 5,622,824, 5,872,003 and 5,691,141 describe methods, systems and kits for exonuclease-mediated mass spectrometric sequencing.

Thus, there is a need for a method for bioagent detection and identification which is both specific and rapid, and in which no nucleic acid sequencing is required. The present invention addresses this need.

**SUMMARY OF THE INVENTION**

The present invention is directed to methods comprising transmitting a sample suspected of containing a bioagent from a remote location to a central location, and analyzing the sample in the central location to confirm the presence or absence of the bioagent. In some embodiments, information regarding the presence or absence of the bioagent is transmitted to the remote location after analysis in the central location. Such information regarding the presence or absence of the bioagent includes the name of the bioagent, a containment protocol for the bioagent, a treatment protocol for the bioagent, or any combination thereof. In some embodiments, the transmitting is carried out using a local area network (LAN) or a wide area network (WAN). In some embodiments, the sample is obtained from a human by a medical personnel or is an inanimate object suspected of being contaminated. In some embodiments, the sample is transmitted in a biohazard container. In some embodiments, the analysis of the bioagent comprises contacting nucleic acid from the bioagent in the sample with at least one pair of oligonucleotide primers that hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, producing an amplification product of the variable nucleic acid sequence, determining a first molecular mass or base composition of the amplification product, and comparing the first molecular mass or base composition to the molecular masses or base compositions of known bioagents, thereby identifying the unknown bioagent in the sample.

The present invention is also directed to methods comprising detecting a bioagent in a first remote location, transmitting the detection status of the bioagent in the first remote location to a centralized location, and transmitting the detection status of the bioagent in the first remote location from the centralized location to at least one additional remote location. In some embodiments, the first remote location is a medical station. In some embodiments, the transmitting is carried out using a local area network (LAN) or a wide area network (WAN). In some embodiments, the detecting the bioagent comprises contacting nucleic acid from the bioagent in the sample with at least one pair of oligonucleotide primers that hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, producing an amplification product of the variable nucleic acid sequence, determining a first molecular mass or base composition of the amplification product, and comparing the first molecular mass or base composition to the molecular masses or base compositions of known bioagents, thereby identifying the unknown bioagent in the sample.



One embodiment of the present invention is directed to methods of determining the presence or absence of a bioagent in a water tower. A sample of water from the tower suspected of containing nucleic acid from a bioagent is contacted with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. The variable nucleic acid sequence is amplified to produce an amplification product. The molecular mass or base composition signature of the amplification product is determined. The molecular mass or base composition signature of the amplification product is compared to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known bioagents, wherein a match indicates the presence of a known bioagent and no match indicates the absence of no bioagent.

Another embodiment of the present invention is directed to methods of identifying an unknown bioagent in a water tower. A sample of water from the tower suspected of containing nucleic acid from a bioagent is contacted with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. The variable nucleic acid sequence is amplified to produce an amplification product. The molecular mass or base composition signature of the amplification product is determined. The molecular mass or base composition signature of the amplification product is compared to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known bioagents, wherein a match identifies the unknown bioagent.

Another embodiment of the present invention is directed to methods of providing information regarding the safety of water in a water tower. A sample of water from the water tower is received. The presence or absence of a harmful bioagent in the sample is determined. Information regarding the safety of the water is transmitted, whereby the presence of a harmful bioagent indicates that the water is unsafe and the absence of a harmful bioagent indicates that the water is safe. Detection of the presence or absence of a harmful bioagent in

the sample can be carried out by contacting the sample of water from the tower suspected of containing nucleic acid from a bioagent with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. The variable nucleic acid sequence is amplified to produce an amplification product. The molecular mass or base composition signature of the amplification product is determined. The molecular mass or base composition signature of the amplification product is compared to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known bioagents, wherein a match identifies the unknown bioagent.

Another embodiment of the present invention is directed to methods of identifying a bioagent present in a water tower using a database of molecular masses or base composition signatures of known bioagents. Nucleic acid from the bioagent is contacted with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. An amplification product of the variable nucleic acid sequence is produced. A first molecular mass or base composition signature of the amplification product is determined. The first molecular mass or base composition signature is compared to the molecular masses or base composition signatures of known bioagents in the database, thereby identifying the bioagent in the water tower.

Another embodiment of the present invention is directed to methods of providing information regarding the bioagent status of a sample. A sample is received and the presence or absence of a bioagent in the sample is detected. Information regarding the bioagent status of the sample is transmitted. In addition, a kit for obtaining a sample may also be provided.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-1H and Figure 2 are consensus diagrams that show examples of conserved regions from 16S rRNA (Fig. 1A-1, 1A-2, 1A-3, 1A-4, and 1A-5), 23S rRNA (3'-half, Fig. 1B, 1C, and 1D; 5'-half, Fig. 1E-F), 23S rRNA Domain I (Fig. 1G), 23S rRNA

Domain IV (Fig. 1H) and 16S rRNA Domain III (Fig. 2) which are suitable for use in the present invention. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. Bases in capital letters are greater than 95% conserved; bases in lower case letters are 90-95% conserved, filled circles are 80-90% conserved; and open circles are less than 80% conserved. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram. The nucleotide sequence of the 16S rRNA consensus sequence is SEQ ID NO:3 and the nucleotide sequence of the 23S rRNA consensus sequence is SEQ ID NO:4.

Figure 2 shows a typical primer amplified region from the 16S rRNA Domain III shown in Figure 1A-1.

Figure 3 is a schematic diagram showing conserved regions in RNase P. Bases in capital letters are greater than 90% conserved; bases in lower case letters are 80-90% conserved; filled circles designate bases which are 70-80% conserved; and open circles designate bases that are less than 70% conserved.

Figure 4 is a schematic diagram of base composition signature determination using nucleotide analog "tags" to determine base composition signatures.

Figure 5 shows the deconvoluted mass spectra of a *Bacillus anthracis* region with and without the mass tag phosphorothioate A (A\*). The two spectra differ in that the measured molecular weight of the mass tag-containing sequence is greater than the unmodified sequence.

Figure 6 shows base composition signature (BCS) spectra from PCR products from *Staphylococcus aureus* (*S. aureus* 16S\_1337F) and *Bacillus anthracis* (*B. anthr.* 16S\_1337F), amplified using the same primers. The two strands differ by only two (AT-->CG) substitutions and are clearly distinguished on the basis of their BCS.

Figure 7 shows that a single difference between two sequences (A14 in *B. anthracis* vs. A15 in *B. cereus*) can be easily detected using ESI-TOF mass spectrometry.

Figure 8 is an ESI-TOF of *Bacillus anthracis* spore coat protein sspE 56mer plus calibrant. The signals unambiguously identify *B. anthracis* versus other *Bacillus* species.

Figure 9 is an ESI-TOF of a *B. anthracis* synthetic 16S\_1228 duplex (reverse and forward strands). The technique easily distinguishes between the forward and reverse strands.

Figure 10 is an ESI-FTICR-MS of a synthetic *B. anthracis* 16S\_1337 46 base pair duplex.

Figure 11 is an ESI-TOF-MS of a 56mer oligonucleotide (3 scans) from the *B. anthracis* saspB gene with an internal mass standard. The internal mass standards are designated by asterisks.

Figure 12 is an ESI-TOF-MS of an internal standard with 5 mM TBA-TFA buffer showing that charge stripping with tributylammonium trifluoroacetate reduces the most abundant charge state from  $[M-8H+]^{8-}$  to  $[M-3H+]^{3-}$ .

Figure 13 is a portion of a secondary structure defining database according to one embodiment of the present invention, where two examples of selected sequences are displayed graphically thereunder.

Figure 14 is a three dimensional graph demonstrating the grouping of sample molecular weight according to species.

Figure 15 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus and mammal infected.

Figure 16 is a three dimensional graph demonstrating the grouping of sample molecular weights according to species of virus, and animal-origin of infectious agent.

Figure 17 is a figure depicting how the triangulation method of the present invention provides for the identification of an unknown bioagent without prior knowledge of the unknown agent. The use of different primer sets to distinguish and identify the unknown is also depicted as primer sets I, II and III within this figure. A three dimensional graph depicts all of bioagent space (170), including the unknown bioagent, which after use of primer set I (171) according to a method according to the present invention further differentiates and classifies bioagents according to major classifications (176) which, upon further analysis using primer set II (172) differentiates the unknown agent (177) from other, known agents (173) and finally, the use of a third primer set (175) further specifies subgroups within the family of the unknown (174).

## DESCRIPTION OF EMBODIMENTS

The present invention provides a combination of a non-PCR biomass detection mode, preferably high-resolution MS, with PCR-based BCS technology using “intelligent primers” which hybridize to conserved sequence regions of nucleic acids derived from a bioagent and which bracket variable sequence regions that uniquely identify the bioagent. The high-resolution MS technique is used to determine the molecular mass and base composition signature (BCS) of the amplified sequence region. This unique “base

composition signature” (BCS) is then input to a maximum-likelihood detection algorithm for matching against a database of base composition signatures in the same amplified region. The present method combines PCR-based amplification technology (which provides specificity) and a molecular mass detection mode (which provides speed and does not require nucleic acid sequencing of the amplified target sequence) for bioagent detection and identification.

The present method allows extremely rapid and accurate detection and identification of bioagents compared to existing methods. Furthermore, this rapid detection and identification is possible even when sample material is impure. The method leverages ongoing biomedical research in virulence, pathogenicity, drug resistance and genome sequencing into a method which provides greatly improved sensitivity, specificity and reliability compared to existing methods, with lower rates of false positives. Thus, the methods are useful in a wide variety of fields, including, but not limited to, those fields discussed below.

In some embodiments, the present invention is directed to methods comprising transmitting a sample suspected of containing a bioagent from a remote location to a central location, and analyzing the sample in the central location to confirm the presence or absence of the bioagent. The sample suspected of containing a bioagent can be any one or more of the samples described herein and can contain any one or more of the bioagents described herein. The sample can be obtained from a human by medical personnel. Alternately, the sample can be an inanimate object suspected of being contaminated.

The remote location can be any location where a sample suspected of containing a bioagent is present. In some embodiments, the remote location is a field location, a place of employment, or a medical station, such as a hospital, clinic, office, or the like. The central location is a location or plurality of locations wherein the sample is analyzed for the presence or absence of the suspected bioagent. The sample can be transmitted from the remote location to the central location by a variety of means including, but not limited to, shipment by airplane, car, truck, train, ship, and the like. The sample can be transmitted in a biohazard container.

The analysis of the bioagent can be carried out by any known method of bioagent detection. In some embodiments, the bioagent detection is carried out by the methods disclosed herein. Such methods include, contacting nucleic acid from the bioagent in the sample with at least one pair of oligonucleotide primers that hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent,

producing an amplification product of the variable nucleic acid sequence, determining a first molecular mass or base composition of the amplification product, and comparing the first molecular mass or base composition to the molecular masses or base compositions of known bioagents, thereby identifying the unknown bioagent in the sample.

5           After the analysis has been carried out, information regarding the presence or absence of the bioagent can be transmitted back to the remote location or to additional remote locations. The information regarding the presence or absence of the bioagent includes, but is not limited to, the name of the bioagent(s), a containment protocol for the bioagent(s), a treatment protocol for the bioagent(s), or any combination thereof. The transmission can be  
10       carried out using a local area network (LAN) or a wide area network (WAN).

          In other embodiments, the present invention is directed to methods comprising detecting a bioagent in a first remote location, transmitting the detection status of the bioagent in the first remote location to a centralized location, and transmitting the detection status of the bioagent in the first remote location from the centralized location to at least one  
15       additional remote location. Such methods comprise an alert system for bioagent detection. The sample suspected of containing a bioagent can be any one or more of the samples described herein and can contain any one or more of the bioagents described herein. The sample can be obtained from a human by medical personnel. Alternately, the sample can be an inanimate object suspected of being contaminated.

20           The detection of the bioagent can be carried out by any known method of bioagent detection. In some embodiments, the bioagent detection is carried out by the methods disclosed herein. Such methods include, contacting nucleic acid from the bioagent in the sample with at least one pair of oligonucleotide primers that hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent,  
25       producing an amplification product of the variable nucleic acid sequence, determining a first molecular mass or base composition of the amplification product, and comparing the first molecular mass or base composition to the molecular masses or base compositions of known bioagents, thereby identifying the unknown bioagent in the sample.

          The remote location can be any location where a sample suspected of containing a  
30       bioagent is detected. In some embodiments, the remote location is a medical station, such as a hospital, clinic, office, or the like, or a scientific station, such as a university or any research institute, that has the capability of detecting bioagents. The central location is a location or plurality of locations that receives information regarding the bioagent detection status at the

remote location(s). The information regarding the bioagent detection status includes, but is not limited to, the identity and address of the first remote location, the name of the bioagent(s), the approximate date of infection or contamination, identification of precise sites of contamination or infection, a prophylactic treatment protocol for the bioagent(s), or any combination thereof. The transmission of the bioagent detection status can be carried out using a local area network (LAN) or a wide area network (WAN), or any other means of communication.

Once a bioagent is detected and the bioagent detection status has been transmitted from the first remote location to the centralized location, the centralized location disperses the information to at least one additional remote location. Thus, the centralized location acts as an alert station and provides warning to other remote locations regarding the bioagent detection in the first remote location. The additional remote locations include medical stations throughout the world, governments throughout the world, health organizations throughout the world, and the like. The centralized location can alert all subscribers to the alert system, or can alert only a subset of all of the subscribers to the alert system.

In other embodiments, the present invention is directed to a method for approving a food product for consumption comprising sampling the food product for the presence or absence of a bioagent, and, if the food product is free of harmful pathogens, indicating on the food product that it is safe for consumption. The food products can be any item that is ingested by an animal including, but not limited to, a meat product, a produce product, or a beverage.

The detection of the bioagent, such as a harmful or non-harmful pathogen, can be carried out by any known method of bioagent detection. In some embodiments, the bioagent detection is carried out by the methods disclosed herein. Such methods include, contacting nucleic acid from the bioagent in the sample with at least one pair of oligonucleotide primers that hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, producing an amplification product of the variable nucleic acid sequence, determining a first molecular mass or base composition of the amplification product, and comparing the first molecular mass or base composition to the molecular masses or base compositions of known bioagents, thereby identifying the unknown bioagent in the sample.

After sampling and testing of the food product, if the food product is free of any harmful pathogens, the food product is indicated as being safe for consumption. Such

indication includes, for example, a label affixed to the packaging material for the food product indicating that the food product has been tested and is free of pathogens.

In some embodiments of the invention, the methods disclosed herein can be used for environmental testing. Detection and discrimination of pathogenic vs. non-pathogenic bacteria, viruses, parasites, fungi and the like, in samples of water, land, air, or other samples, can be carried out. Water samples can be obtained from, for example, lakes, rivers, oceans, streams, water treatment systems, rainwater, groundwater, water table, reservoirs, wells, bottled water, and the like. Air samples can be obtained from ventilation systems, airplane cabins, schools, hospitals, mass transit locations such as subways, train stations, airports, and the like. Conditions such as sick building syndrome can be detected. Land samples can be obtained from any location.

In other embodiments of the invention, the methods disclosed herein can be used to screen blood and other bodily fluids and tissues for pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like. Animal samples, including but not limited to, blood and other bodily fluid and tissue samples, can be obtained from living animals, who are either known or not known to or suspected of having a disease, infection, or condition. Alternately, animal samples such as blood and other bodily fluid and tissue samples can be obtained from deceased animals. Blood samples can be further separated into plasma or cellular fractions and further screened as desired. Bodily fluids and tissues can be obtained from any part of the animal or human body. Animal samples can be obtained from, for example, mammals and humans.

In other embodiments of the invention, the methods disclosed herein can be used for forensics. For example, medical examiners can use the present invention to determine the cause of death. In addition, epidemiologists, for example, can use the present methods to determine the geographic origin of a particular strain of bacteria or virus. For example, a particular strain of bacteria or virus may have a sequence difference that is associated with a particular area of a country or the world and identification of such a sequence difference can lead to the identification of the geographic origin and epidemiological tracking of the spread of the particular disease, disorder or condition associated with the detected virus or bacteria. In addition, carriers of particular DNA or diseases, such as mammals, non-mammals, birds, insects, and plants, can be tracked by screening SNPs, VNTRs, or polyA, for example. Diseases, such as malaria, can be tracked by screening commensals, such as mosquitos.



In other embodiments of the invention, the methods disclosed herein can be used for detecting the presence of pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like in samples of foodstuff or cosmetics. For example, food and wine can be examined for the presence of pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like. Particular types of foods susceptible to bioagent contamination, such as agricultural products, meat products and eggs, can be examined for pathogenic organisms such as *E. coli* and *Salmonella* species. Such examination procedures can be used by, for example, the wholesalers of foodstuffs and beverages, or by regulatory agencies such as the U.S. Department of Agriculture and the Food and Drug Administration. In addition, grapes and wines, for example, can be examined using the present methods to detect particular strains of bacteria or yeast that may indicate a particular time upon which to harvest the grapes or alter the wine-making process.

In other embodiments of the invention, the methods disclosed herein can be used for detecting the presence of bioagents in a container, such as a package, box, envelope, mail tube, railroad box car, and the like. For example, mail and package delivery entities and agencies, both domestic and abroad, as well investigative agencies such as the FBI and ATF can use the present methods to detect bioagents in containers.

In other embodiments of the invention, the methods disclosed herein can be used for detecting the presence of pathogenic and non-pathogenic bacteria, viruses, parasites, fungi and the like in organ donors and/or in organs from donors. Such examination can result in the prevention of the transfer of, for example, viruses such as West Nile virus, hepatitis viruses, human immunodeficiency virus, and the like from a donor to a recipient via a transplanted organ. The methods disclosed herein can also be used for detection of host versus graft or graft versus host rejection issues related to organ donors by detecting the presence of particular antigens in either the graft or host known or suspected of causing such rejection. In particular, the bioagents in this regard are the antigens of the major histocompatibility complex, such as the HLA antigens.

In other embodiments of the invention, the methods disclosed herein can be used for detection and identification of livestock infections such as, for example, mad cow disease, hoof and mouth disease, and the like. Livestock includes, but is not limited to, cows, pigs, sheep, chickens, turkeys, goats, and other farm animals.

In other embodiments of the invention, the methods disclosed herein can be used for pharmacogenetic analysis and medical diagnosis including, but not limited to, cancer

diagnosis based on mutations and polymorphisms, drug resistance and susceptibility testing, screening for and/or diagnosis of genetic diseases and conditions, and diagnosis of infectious diseases and conditions. The present methods can also be used to detect and track emerging infectious diseases, such as West Nile virus infection, mad cow disease, and HIV-related diseases.

The present methods can be used to detect and classify any biological agent, including bacteria, viruses, fungi and toxins. As one example, where the agent is a biological threat, the information obtained is used to determine practical information needed for countermeasures, including toxin genes, pathogenicity islands and antibiotic resistance genes.

In addition, the methods can be used to identify natural or deliberate engineering events including chromosome fragment swapping, molecular breeding (gene shuffling) and emerging infectious diseases.

One embodiment of the present invention is directed to methods of determining the presence or absence of a bioagent in a water tower. The water tower can be any container containing water. The water tower can be used for heating and/or cooling and can service a residence, commercial, or industrial building. The water tower can also be in the form of a waste- water pond. The water pond may be "suspected" of containing a bioagent. As used herein, the term "suspected" simply means that the water tower may contain a bioagent. One skilled in the art may have a particular reason to suspect that the water tower indeed harbors a bioagent. Alternately, one skilled in the art may have no particular reason to suspect that the water tower harbors a bioagent but, rather, is interested in testing the water in the water tower.

A sample of water from the tower suspected of containing nucleic acid from a bioagent is contacted with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. The variable nucleic acid sequence is amplified to produce an amplification product. The molecular mass or base composition signature of the amplification product is determined. The molecular mass or base composition signature of the amplification product is compared to one or more molecular masses or base composition signatures of amplification products obtained from a

plurality of known bioagents, wherein a match indicates the presence of a known bioagent and no match indicates the absence of no bioagent.

Another embodiment of the present invention is directed to methods of identifying an unknown bioagent in a water tower. A sample of water from the tower suspected of containing nucleic acid from a bioagent is contacted with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. The variable nucleic acid sequence is amplified to produce an amplification product. The molecular mass or base composition signature of the amplification product is determined. The molecular mass or base composition signature of the amplification product is compared to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known bioagents, wherein a match identifies the unknown bioagent.

Another embodiment of the present invention is directed to methods of providing information regarding the safety of water in a water tower. A sample of water from the water tower is received. The sample of water can be obtained from the water tower by an owner of the water tower, a caretaker of the water tower, or by a person servicing the water tower, or by anyone associated with the water tower. The sample of water is received, at least eventually, by a party capable of carrying out methods of detecting bioagents. The presence or absence of a harmful bioagent in the sample is determined. One skilled in the art is capable of determining which bioagents are harmful to animals, and in particular, humans. Information regarding the safety of the water is transmitted back to an owner of the water tower, a caretaker of the water tower, or by a person servicing the water tower, or by anyone associated with the water tower. Such information can also be transmitted to any government or private agency responsible for water quality standards, as well as any government or private agency responsible for human health. The presence of a harmful bioagent indicates that the water is unsafe and the absence of a harmful bioagent indicates that the water is safe.

Detection of the presence or absence of a harmful bioagent in the sample can be carried out by contacting the sample of water from the tower suspected of containing nucleic acid from a bioagent with either at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid

sequence of the bioagent, or with at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid. The variable nucleic acid sequence is amplified to produce an amplification product. The molecular mass or base composition signature of the amplification product is determined. The molecular mass or base composition signature of the amplification product is compared to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known bioagents, wherein a match identifies the unknown bioagent.

Another embodiment of the present invention is directed to methods of providing information regarding the bioagent status of a sample. The bioagent status of a sample includes, but is not limited to, whether or not a bioagent is present in the sample, and/or the identity of the bioagent. A sample includes any portion or amount of any material known or suspected of containing or being associated with or capable of supporting a bioagent, including an air sample, water sample, land sample, and the like. The sample can also be a clothing sample, food sample, beverage sample, and the like. Samples can also include water from a swimming pool. A kit for obtaining a sample may also be provided. The kits can contain sampling devices well known in the art for obtaining various types of sample, depending on their form (i.e., solid, liquid, or gas). The sample is received and the presence or absence of a bioagent in the sample is detected by any methodology, including the methods of detection and/or identification disclosed herein. Information regarding the bioagent status of the sample is transmitted to the requestor, to a private or governmental agency or organization responsible for public safety, or to a private or governmental agency or organization responsible for environmental safety.

For example, a business entity may provide a kit for obtaining water samples from a swimming pool. Owners, operators, and users of a swimming pool, either private or commercial, may desire the health status of the swimming pool. The sample is forwarded to the business entity so that any bioagents in the sample can be detected. Upon detecting the presence or absence of a bioagent in the sample, the business entity can transmit information regarding the bioagent status back to the requestor. Such information can include whether there is any bioagent in the sample, the identity of the bioagent in the sample, health risk(s) associated with the bioagent, and methods of reducing levels of or eliminating the bioagent, or any combination thereof.

The embodiments of the present invention provide a combination of a non-PCR biomass detection mode, preferably high-resolution MS, with PCR-based BCS or mass technology using “intelligent primers” that hybridize to conserved sequence regions of nucleic acids derived from a bioagent and that bracket variable sequence regions that uniquely identify the bioagent. Alternately, the variable nucleic acid sequence of the bioagent is flanked by a primer and a natural stop region of the nucleic acid. Natural stop regions of a nucleic acid include, but are not limited to, the natural stop codons present in the genetic code of the organism from which the bioagent is derived, as well as any other stop regions. The high-resolution MS technique is used to determine the molecular mass and/or base composition signature (BCS) of the amplified sequence region. This unique mass or unique BCS is then input to a maximum-likelihood detection algorithm for matching against a database of masses or BCSs in the same amplified region. The present method combines PCR-based amplification technology (which provides specificity) and a molecular mass detection mode (which provides speed and does not require nucleic acid sequencing of the amplified target sequence) for bioagent detection and identification.

In some embodiments of the invention, the nucleic acid from the bioagent (e.g., plant cells, animal cells, bacterial cells, viruses, parasites, fungi, molds, toxin genes, and bioregulating compound) is a nucleic acid present in numerous species. Bacteria have a common set of absolutely required genes. About 250 genes are present in all bacterial species (*Proc. Natl. Acad. Sci. U.S.A.*, 1996, 93, 10268; *Science*, 1995, 270, 397), including tiny genomes like *Mycoplasma*, *Ureaplasma* and *Rickettsia*. These genes encode proteins involved in translation, replication, recombination and repair, transcription, nucleotide metabolism, amino acid metabolism, lipid metabolism, energy generation, uptake, secretion and the like. Examples of these proteins are DNA polymerase III beta, elongation factor TU, heat shock protein groEL, RNA polymerase beta, phosphoglycerate kinase, NADH dehydrogenase, DNA ligase, DNA topoisomerase and elongation factor G. Operons can also be targeted using the present method. One example of an operon is the bfp operon from enteropathogenic *E. coli*. Multiple core chromosomal genes can be used to classify bacteria at a genus or genus species level to determine if an organism has threat potential. The methods can also be used to detect pathogenicity markers (plasmid or chromosomal) and antibiotic resistance genes to confirm the threat potential of an organism and to direct countermeasures.

A theoretically ideal bioagent detector would identify, quantify, and report the

complete nucleic acid sequence of every bioagent that reached the sensor. The complete sequence of the nucleic acid component of a pathogen would provide all relevant information about the threat, including its identity and the presence of drug-resistance or pathogenicity markers. This ideal has not yet been achieved. However, the present invention provides a straightforward strategy for obtaining information with the same practical value using base composition signatures (BCS). While the base composition of a gene fragment is not as information-rich as the sequence itself, there is no need to analyze the complete sequence of the gene if the short analyte sequence fragment is properly chosen. A database of reference sequences can be prepared in which each sequence is indexed to a unique base composition signature, so that the presence of the sequence can be inferred with accuracy from the presence of the signature. The advantage of base composition signatures is that they can be quantitatively measured in a massively parallel fashion using multiplex PCR (PCR in which two or more primer pairs amplify target sequences simultaneously) and mass spectrometry. These multiple primer amplified regions uniquely identify most threat and ubiquitous background bacteria and viruses. In addition, cluster-specific primer pairs distinguish important local clusters (e.g., anthracis group).

In the context of this invention, a "bioagent" is any organism, living or dead, or a nucleic acid derived from such an organism. Examples of bioagents include but are not limited to cells (including but not limited to human clinical samples, bacterial cells and other pathogens) viruses, parasites, fungi, toxin genes and bioregulating compounds. Samples may be alive or dead or in a vegetative state (for example, vegetative bacteria or spores) and may be encapsulated or bioengineered.

As used herein, a "base composition signature" (BCS) is the exact base composition from selected fragments of nucleic acid sequences that uniquely identifies the target gene and source organism. BCS can be thought of as unique indexes of specific genes.

As used herein, "intelligent primers" are primers that bind to sequence regions that flank an intervening variable region. In one embodiment, these sequence regions that flank the variable region are highly conserved among different species of bioagent. For example, the sequence regions may be highly conserved among all *Bacillus* species. By the term "highly conserved", it is meant that the sequence regions exhibit between about 80-100%, more preferably between about 90-100% and most preferably between about 95-100% identity. Percent identity can be determined by, for example, the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University

Research Park, Madison WI), using the default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489, which is incorporated herein by reference in its entirety). Examples of intelligent primers that amplify regions of the 16S and 23S rRNA are shown in Figures 1A-1I. A typical primer amplified region in 16S rRNA is shown in Figure 2. The arrows represent primers that bind to highly conserved regions that flank a variable region in 16S rRNA domain III. The amplified region is the stem-loop structure under "1100-1188."

Alternately, a variable nucleic acid sequence of the bioagent is flanked by the intelligent primer and a natural stop region of the nucleic acid. Natural stop regions of a nucleic acid include, but are not limited to, the natural stop codons present in the genetic code of the organism from which the bioagent is derived, as well as any other stop regions.

The variable nucleic acid sequence can uniquely identify the bioagent as indigenous to a particular geographic location of the world. The variable nucleic acid sequence can be the portion of a nucleic acid molecule that distinguishes one strain of bacteria or virus which is indigenous to one particular geographic location of the world from another strain of bacteria or virus that is indigenous to a different geographic location of the world. In addition, the variable nucleic acid sequence can be the portion of a nucleic acid molecule that distinguishes one strain of animal or plant cell from another strain of plant or animal cell. Likewise for fungi, mold, and parasites. Thus, it is the variable nucleic acid sequence that can link a particular bioagent to a particular geographical region of the world.

One main advantage of the detection methods of the present invention is that the primers need not be specific for a particular bacterial species, or even genus, such as *Bacillus* or *Streptomyces*. Instead, the primers recognize highly conserved regions across hundreds of bacterial species including, but not limited to, the species described herein. Thus, the same primer pair can be used to identify any desired bacterium because it will bind to the conserved regions that flank a variable region specific to a single species, or common to several bacterial species, allowing nucleic acid amplification of the intervening sequence and determination of its molecular weight and base composition. For example, the 16S\_971-1062, 16S\_1228-1310 and 16S\_1100-1188 regions are 98-99% conserved in about 900 species of bacteria (16S=16S rRNA, numbers indicate nucleotide position). In one embodiment of the present invention, primers used in the present method bind to one or more of these regions or portions thereof.

The present invention provides a combination of a non-PCR biomass detection mode, preferably high-resolution MS, with nucleic acid amplification-based BCS technology using “intelligent primers” which hybridize to conserved regions and which bracket variable regions that uniquely identify the bioagent(s). Although the use of PCR is preferred, other nucleic acid amplification techniques may also be used, including ligase chain reaction (LCR) and strand displacement amplification (SDA). The high-resolution MS technique allows separation of bioagent spectral lines from background spectral lines in highly cluttered environments. The resolved spectral lines are then translated to BCS which are input to a maximum-likelihood detection algorithm matched against spectra for one or more known BCS. Preferably, the bioagent BCS spectrum is matched against one or more databases of BCS from vast numbers of bioagents. Preferably, the matching is done using a maximum-likelihood detection algorithm.

In one embodiment, base composition signatures are quantitatively measured in a massively parallel fashion using the polymerase chain reaction (PCR), preferably multiplex PCR, and mass spectrometric (MS) methods. Sufficient quantities of nucleic acids should be present for detection of bioagents by MS. A wide variety of techniques for preparing large amounts of purified nucleic acids or fragments thereof are well known to those of skill in the art. PCR requires one or more pairs of oligonucleotide primers that bind to regions which flank the target sequence(s) to be amplified. These primers prime synthesis of a different strand of DNA, with synthesis occurring in the direction of one primer towards the other primer. The primers, DNA to be amplified, a thermostable DNA polymerase (e.g. *Taq* polymerase), the four deoxynucleotide triphosphates, and a buffer are combined to initiate DNA synthesis. The solution is denatured by heating, then cooled to allow annealing of newly added primer, followed by another round of DNA synthesis. This process is typically repeated for about 30 cycles, resulting in amplification of the target sequence.

The “intelligent primers” define the target sequence region to be amplified and analyzed. In one embodiment, the target sequence is a ribosomal RNA (rRNA) gene sequence. With the complete sequences of many of the smallest microbial genomes now available, it is possible to identify a set of genes that defines “minimal life” and identify composition signatures that uniquely identify each gene and organism. Genes that encode core life functions such as DNA replication, transcription, ribosome structure, translation, and transport are distributed broadly in the bacterial genome and are preferred regions for BCS analysis. Ribosomal RNA (rRNA) genes comprise regions that provide useful base



composition signatures. Like many genes involved in core life functions, rRNA genes contain sequences that are extraordinarily conserved across bacterial domains interspersed with regions of high variability that are more specific to each species. The variable regions can be utilized to build a database of base composition signatures. The strategy involves creating a structure-based alignment of sequences of the small (16S) and the large (23S) subunits of the rRNA genes. For example, there are currently over 13,000 sequences in the ribosomal RNA database that has been created and maintained by Robin Gutell, University of Texas at Austin, and is publicly available on the Institute for Cellular and Molecular Biology web page on the world wide web of the Internet at, for example, "rna.icmb.utexas.edu/." There is also a publicly available rRNA database created and maintained by the University of Antwerp, Belgium on the world wide web of the Internet at, for example, "rna.uia.ac.be."

These databases have been analyzed to determine regions that are useful as base composition signatures. The characteristics of such regions include: a) between about 80 and 100%, preferably > about 95% identity among species of the particular bioagent of interest, of upstream and downstream nucleotide sequences which serve as sequence amplification primer sites; b) an intervening variable region which exhibits no greater than about 5% identity among species; and c) a separation of between about 30 and 1000 nucleotides, preferably no more than about 50-250 nucleotides, and more preferably no more than about 60-100 nucleotides, between the conserved regions.

Due to their overall conservation, the flanking rRNA primer sequences serve as good "universal" primer binding sites to amplify the region of interest for most, if not all, bacterial species. The intervening region between the sets of primers varies in length and/or composition, and thus provides a unique base composition signature.

Examples of intelligent primers that amplify regions of the 16S and 23S rRNA are shown in Figures 1A-1H. A typical primer amplified region in 16S rRNA is shown in Figure 2. The arrows represent primers that bind to highly conserved regions which flank a variable region in 16S rRNA domain III. The amplified region is the stem-loop structure under "1100-1188." It is advantageous to design the broad range survey intelligent primers to minimize the number of primers required for the analysis, and to allow detection of multiple members of a bioagent division using a single pair of primers. The advantage of using broad range survey intelligent primers is that once a bioagent is broadly identified, the process of further identification at species and sub-species levels is facilitated by directing the choice of additional intelligent primers.

Identification of bioagents can be accomplished at different levels using intelligent primers suited to resolution of each individual level of identification. "Broad range survey" intelligent primers are designed with the objective of identifying a bioagent as a member of a particular division of bioagents. A "bioagent division" is defined as group of bioagents above the species level and includes but is not limited to: orders, families, classes, clades, genera or other such groupings of bioagents above the species level. As a non-limiting example, members of the *Bacillus/Clostridia* group or gamma-proteobacteria group may be identified as such by employing broad range survey intelligent primers such as primers which target 16S or 23S ribosomal RNA.

In some embodiments, broad range survey intelligent primers are capable of identification of bioagents at the species level. One main advantage of the detection methods of the present invention is that the broad range survey intelligent primers need not be specific for a particular bacterial species, or even genus, such as *Bacillus* or *Streptomyces*. Instead, the primers recognize highly conserved regions across hundreds of bacterial species including, but not limited to, the species described herein. Thus, the same broad range survey intelligent primer pair can be used to identify any desired bacterium because it will bind to the conserved regions that flank a variable region specific to a single species, or common to several bacterial species, allowing unbiased nucleic acid amplification of the intervening sequence and determination of its molecular weight and base composition. For example, the 16S\_971-1062, 16S\_1228-1310 and 16S\_1100-1188 regions are 98-99% conserved in about 900 species of bacteria (16S=16S rRNA, numbers indicate nucleotide position). In one embodiment of the present invention, primers used in the present method bind to one or more of these regions or portions thereof.

"Division-wide" intelligent primers are designed with an objective of identifying a bioagent at the species level. As a non-limiting example, a *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis* can be distinguished from each other using division-wide intelligent primers. Division-wide intelligent primers are not always required for identification at the species level because broad range survey intelligent primers may provide sufficient identification resolution to accomplishing this identification objective.

"Drill-down" intelligent primers are designed with an objective of identifying a sub-species characteristic of a bioagent. A "sub-species characteristic" is defined as a property imparted to a bioagent at the sub-species level of identification as a result of the presence or absence of a particular segment of nucleic acid. Such sub-species characteristics include, but

are not limited to, strains, sub-types, pathogenicity markers such as antibiotic resistance genes, pathogenicity islands, toxin genes and virulence factors. Identification of such sub-species characteristics is often critical for determining proper clinical treatment of pathogen infections.

5 It is advantageous to design the "intelligent primers" to be as universal as possible to minimize the number of primers which need to be synthesized, and to allow detection of multiple species using a single pair of primers. These primer pairs can be used to amplify variable regions in these species. Because any variation (due to codon wobble in the 3<sup>rd</sup> position) in these conserved regions among species is likely to occur in the third position of a  
10 DNA triplet, oligonucleotide primers can be designed such that the nucleotide corresponding to this position is a base which can bind to more than one nucleotide, referred to herein as a "universal base." For example, under this "wobble" pairing, inosine (I) binds to U, C or A; guanine (G) binds to U or C, and uridine (U) binds to U or C. Other examples of universal bases include nitroindoles such as 5-nitroindole or 3-nitropyrrole (Loakes *et al.*, *Nucleosides and Nucleotides*, 1995, 14, 1001-1003), the degenerate nucleotides dP or dK (Hill *et al.*), an  
15 acyclic nucleoside analog containing 5-nitroindazole (Van Aerschot *et al.*, *Nucleosides and Nucleotides*, 1995, 14, 1053-1056) or the purine analog 1-(2-deoxy- $\beta$ -D-ribofuranosyl)-imidazole-4-carboxamide (Sala *et al.*, *Nucl. Acids Res.*, 1996, 24, 3302-3306).

In another embodiment of the invention, to compensate for the somewhat weaker  
20 binding by the "wobble" base, the oligonucleotide primers are designed such that the first and second positions of each triplet are occupied by nucleotide analogs which bind with greater affinity than the unmodified nucleotide. Examples of these analogs include, but are not limited to, 2,6-diaminopurine which binds to thymine, propyne T which binds to adenine and propyne C and phenoxazines, including G-clamp, which binds to G. Propynes are described  
25 in U.S. Patent Nos. 5,645,985, 5,830,653 and 5,484,908, each of which is incorporated herein by reference in its entirety. Phenoxazines are described in U.S. Patent Nos. 5,502,177, 5,763,588, and 6,005,096, each of which is incorporated herein by reference in its entirety. G-clamps are described in U.S. Patent Nos. 6,007,992 and 6,028,183, each of which is incorporated herein by reference in its entirety.

30 Bacterial biological warfare agents capable of being detected by the present methods include, but are not limited to, *Bacillus anthracis* (anthrax), *Yersinia pestis* (pneumonic plague), *Franciscella tularensis* (tularemia), *Brucella suis*, *Brucella abortus*, *Brucella melitensis* (undulant fever), *Burkholderia mallei* (glanders), *Burkholderia pseudomallei*

(melioidosis), *Salmonella typhi* (typhoid fever), *Rickettsia typhi* (epidemic typhus), *Rickettsia prowasekii* (endemic typhus) and *Coxiella burnetii* (Q fever), *Rhodobacter capsulatus*, *Chlamydia pneumoniae*, *Escherichia coli*, *Shigella dysenteriae*, *Shigella flexneri*, *Bacillus cereus*, *Clostridium botulinum*, *Coxiella burnetii*, *Pseudomonas aeruginosa*,  
5 *Legionella pneumophila*, and *Vibrio cholerae*.

Besides 16S and 23S rRNA, other target regions suitable for use in the present invention for detection of bacteria include, but are not limited to, 5S rRNA and RNase P (Figure 3).

Biological warfare fungus biowarfare agents include, but are not limited to,  
10 *coccidioides immitis* (Coccidioidomycosis).

Biological warfare toxin genes capable of being detected by the methods of the present invention include, but are not limited to, botulism, T-2 mycotoxins, ricin, staph enterotoxin B, shigatoxin, abrin, aflatoxin, *Clostridium perfringens* epsilon toxin, conotoxins, diacetoxyscirpenol, tetrodotoxin and saxitoxin.

15 Biological warfare viral threat agents are mostly RNA viruses (positive-strand and negative-strand), with the exception of smallpox. Every RNA virus is a family of related viruses (quasispecies). These viruses mutate rapidly and the potential for engineered strains (natural or deliberate) is very high. RNA viruses cluster into families that have conserved RNA structural domains on the viral genome (e.g., virion components, accessory proteins)  
20 and conserved housekeeping genes that encode core viral proteins including, for single strand positive strand RNA viruses, RNA-dependent RNA polymerase, double stranded RNA helicase, chymotrypsin-like and papain-like proteases and methyltransferases.

Examples of (-)-strand RNA viruses include, but are not limited to, arenaviruses (e.g., sabia virus, lassa fever, Machupo, Argentine hemorrhagic fever, flexal virus),  
25 bunyaviruses (e.g., hantavirus, nairovirus, phlebovirus, hantaan virus, Congo-crimean hemorrhagic fever, rift valley fever), and mononegavirales (e.g., filovirus, paramyxovirus, ebola virus, Marburg, equine morbillivirus).

Examples of (+)-strand RNA viruses include, but are not limited to, picornaviruses (e.g., coxsackievirus, echovirus, human coxsackievirus A, human echovirus, human  
30 enterovirus, human poliovirus, hepatitis A virus, human parechovirus, human rhinovirus), astroviruses (e.g., human astrovirus), calciviruses (e.g., chiba virus, chitta virus, human calcivirus, norwalk virus), nidovirales (e.g., human coronavirus, human torovirus), flaviviruses (e.g., dengue virus 1-4, Japanese encephalitis virus, Kyanasur forest disease

virus, Murray Valley encephalitis virus, Rocio virus, St. Louis encephalitis virus, West Nile virus, yellow fever virus, hepatitis c virus) and togaviruses (e.g., Chikugunya virus, Eastern equine encephalitis virus, Mayaro virus, O'nyong-nyong virus, Ross River virus, Venezuelan equine encephalitis virus, Rubella virus, hepatitis E virus). The hepatitis C virus has a 5'-  
5 untranslated region of 340 nucleotides, an open reading frame encoding 9 proteins having 3010 amino acids and a 3'-untranslated region of 240 nucleotides. The 5'-UTR and 3'-UTR are 99% conserved in hepatitis C viruses.

In one embodiment, the target gene is an RNA-dependent RNA polymerase or a helicase encoded by (+)-strand RNA viruses, or RNA polymerase from a (-)-strand RNA  
10 virus. (+)-strand RNA viruses are double stranded RNA and replicate by RNA-directed RNA synthesis using RNA-dependent RNA polymerase and the positive strand as a template. Helicase unwinds the RNA duplex to allow replication of the single stranded RNA. These viruses include viruses from the family picornaviridae (e.g., poliovirus, coxsackievirus, echovirus), togaviridae (e.g., alphavirus, flavivirus, rubivirus), arenaviridae (e.g.,  
15 lymphocytic choriomeningitis virus, lassa fever virus), cononaviridae (e.g., human respiratory virus) and Hepatitis A virus. The genes encoding these proteins comprise variable and highly conserved regions which flank the variable regions.

In some embodiments, the bioagent is one that is capable and/or likely to grow and/or exist in a water environment. Such bioagents include, but are not limited to,  
20 *Salmonella speceis*, *Shigella species*, *Cholera species*, hepatitis virus species, *E. coli*, and *Legionella pneumophila*.

In another embodiment, the detection scheme for the PCR products generated from the bioagent(s) incorporates at least three features. First, the technique simultaneously detects and differentiates multiple (generally about 6-10) PCR products. Second, the technique  
25 provides a BCS that uniquely identifies the bioagent from the possible primer sites. Finally, the detection technique is rapid, allowing multiple PCR reactions to be run in parallel.

In one embodiment, the method can be used to detect the presence of antibiotic resistance and/or toxin genes in a bacterial species. For example, *Bacillus anthracis* comprising a tetracycline resistance plasmid and plasmids encoding one or both anthracis  
30 toxins (px01 and/or px02) can be detected by using antibiotic resistance primer sets and toxin gene primer sets. If the *B. anthracis* is positive for tetracycline resistance, then a different antibiotic, for example quinalone, is used.

Mass spectrometry (MS)-based detection of PCR products provides all of these

features with additional advantages. MS is intrinsically a parallel detection scheme without the need for radioactive or fluorescent labels, since every amplification product with a unique base composition is identified by its molecular mass. The current state of the art in mass spectrometry is such that less than femtomole quantities of material can be readily analyzed to afford information about the molecular contents of the sample. An accurate assessment of the molecular mass of the material can be quickly obtained, irrespective of whether the molecular weight of the sample is several hundred, or in excess of one hundred thousand atomic mass units (amu) or Daltons. Intact molecular ions can be generated from amplification products using one of a variety of ionization techniques to convert the sample to gas phase. These ionization methods include, but are not limited to, electrospray ionization (ES), matrix-assisted laser desorption ionization (MALDI) and fast atom bombardment (FAB). For example, MALDI of nucleic acids, along with examples of matrices for use in MALDI of nucleic acids, are described in WO 98/54751 (Genetrace, Inc.).

In some embodiments, large DNAs and RNAs, or large amplification products therefrom, can be digested with restriction endonucleases prior to ionization. Thus, for example, an amplification product that was 10 kDa could be digested with a series of restriction endonucleases to produce a panel of, for example, 100 Da fragments. Restriction endonucleases and their sites of action are well known to the skilled artisan. In this manner, mass spectrometry can be performed for the purposes of restriction mapping.

Upon ionization, several peaks are observed from one sample due to the formation of ions with different charges. Averaging the multiple readings of molecular mass obtained from a single mass spectrum affords an estimate of molecular mass of the bioagent. Electrospray ionization mass spectrometry (ESI-MS) is particularly useful for very high molecular weight polymers such as proteins and nucleic acids having molecular weights greater than 10 kDa, since it yields a distribution of multiply-charged molecules of the sample without causing a significant amount of fragmentation.

The mass detectors used in the methods of the present invention include, but are not limited to, Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS), ion trap, quadrupole, magnetic sector, time of flight (TOF), Q-TOF, and triple quadrupole.

In general, the mass spectrometric techniques which can be used in the present invention include, but are not limited to, tandem mass spectrometry, infrared multiphoton dissociation and pyrolytic gas chromatography mass spectrometry (PGC-MS). In one embodiment of the invention, the bioagent detection system operates continually in bioagent

detection mode using pyrolytic GC-MS without PCR for rapid detection of increases in biomass (for example, increases in fecal contamination of drinking water or of germ warfare agents). To achieve minimal latency, a continuous sample stream flows directly into the PGC-MS combustion chamber. When an increase in biomass is detected, a PCR process is automatically initiated. Bioagent presence produces elevated levels of large molecular fragments from, for example, about 100-7,000 Da which are observed in the PGC-MS spectrum. The observed mass spectrum is compared to a threshold level and when levels of biomass are determined to exceed a predetermined threshold, the bioagent classification process described hereinabove (combining PCR and MS, preferably FT-ICR MS) is initiated. Optionally, alarms or other processes (halting ventilation flow, physical isolation) are also initiated by this detected biomass level.

The accurate measurement of molecular mass for large DNAs is limited by the adduction of cations from the PCR reaction to each strand, resolution of the isotopic peaks from natural abundance  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes, and assignment of the charge state for any ion.

The cations are removed by in-line dialysis using a flow-through chip that brings the solution containing the PCR products into contact with a solution containing ammonium acetate in the presence of an electric field gradient orthogonal to the flow. The latter two problems are addressed by operating with a resolving power of  $>100,000$  and by incorporating isotopically depleted nucleotide triphosphates into the DNA. The resolving power of the instrument is also a consideration. At a resolving power of 10,000, the modeled signal from the  $[\text{M}-14\text{H}]^{14+}$  charge state of an 84mer PCR product is poorly characterized and assignment of the charge state or exact mass is impossible. At a resolving power of 33,000, the peaks from the individual isotopic components are visible. At a resolving power of 100,000, the isotopic peaks are resolved to the baseline and assignment of the charge state for the ion is straightforward. The  $[^{13}\text{C}, ^{15}\text{N}]$ -depleted triphosphates are obtained, for example, by growing microorganisms on depleted media and harvesting the nucleotides (Batey *et al.*, *Nucl. Acids Res.*, **1992**, *20*, 4515-4523).

While mass measurements of intact nucleic acid regions are believed to be adequate to determine most bioagents, tandem mass spectrometry ( $\text{MS}^n$ ) techniques may provide more definitive information pertaining to molecular identity or sequence. Tandem MS involves the coupled use of two or more stages of mass analysis where both the separation and detection steps are based on mass spectrometry. The first stage is used to select an ion or component of a sample from which further structural information is to be obtained. The selected ion is then

fragmented using, e.g., blackbody irradiation, infrared multiphoton dissociation, or collisional activation. For example, ions generated by electrospray ionization (ESI) can be fragmented using IR multiphoton dissociation. This activation leads to dissociation of glycosidic bonds and the phosphate backbone, producing two series of fragment ions, called the *w*-series (having an intact 3' terminus and a 5' phosphate following internal cleavage) and the *α*-Base series (having an intact 5' terminus and a 3' furan).

The second stage of mass analysis is then used to detect and measure the mass of these resulting fragments of product ions. Such ion selection followed by fragmentation routines can be performed multiple times so as to essentially completely dissect the molecular sequence of a sample.

If there are two or more targets of similar base composition or mass, or if a single amplification reaction results in a product which has the same mass as two or more bioagent reference standards, they can be distinguished by using mass-modifying "tags." In this embodiment of the invention, a nucleotide analog or "tag" is incorporated during amplification (e.g., a 5-(trifluoromethyl) deoxythymidine triphosphate) which has a different molecular weight than the unmodified base so as to improve distinction of masses. Such tags are described in, for example, PCT WO97/33000, which is incorporated herein by reference in its entirety. This further limits the number of possible base compositions consistent with any mass. For example, 5-(trifluoromethyl)deoxythymidine triphosphate can be used in place of dTTP in a separate nucleic acid amplification reaction. Measurement of the mass shift between a conventional amplification product and the tagged product is used to quantitate the number of thymidine nucleotides in each of the single strands. Because the strands are complementary, the number of adenosine nucleotides in each strand is also determined.

In another amplification reaction, the number of G and C residues in each strand is determined using, for example, the cytidine analog 5-methylcytosine (5-meC) or propyne C. The combination of the A/T reaction and G/C reaction, followed by molecular weight determination, provides a unique base composition. This method is summarized in Figure 4 and Table 1.

Table 1

Mass tag	Double strand sequence	Single strand Sequence	Total mass this strand	Base info this strand	Base info other strand	Total base comp. Top strand	Total base comp. Bottom strand
----------	------------------------	------------------------	------------------------	-----------------------	------------------------	-----------------------------	--------------------------------



T* <sub>mass</sub> (T*-T) = x	T*ACGT*ACGT* AT*GCAT*GCA	T*ACGT*ACGT *	3x	3T	3A	3T 2A 2C 2G	3A 2T 2G 2C
		AT*GCAT*GCA	2x	2T	2A		
C* <sub>mass</sub> (C*-C) = y	TAC*GTAC*GT ATGC*ATGC*A	TAC*GTAC*GT	2x	2C	2G		
		ATGC*ATGC*A	2x	2C	2G		

The mass tag phosphorothioate A (A\*) was used to distinguish a *Bacillus anthracis* cluster. The *B. anthracis* (A<sub>14</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) had an average MW of 14072.26, and the *B. anthracis* (A<sub>1</sub>A\*<sub>13</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) had an average molecular weight of 14281.11 and the phosphorothioate A  
5 had an average molecular weight of +16.06 as determined by ESI-TOF MS. The deconvoluted spectra are shown in Figure 5.

In another example, assume the measured molecular masses of each strand are 30,000.115Da and 31,000.115 Da respectively, and the measured number of dT and dA residues are (30,28) and (28,30). If the molecular mass is accurate to 100 ppm, there are 7  
10 possible combinations of dG+dC possible for each strand. However, if the measured molecular mass is accurate to 10 ppm, there are only 2 combinations of dG+dC, and at 1 ppm accuracy there is only one possible base composition for each strand.

Signals from the mass spectrometer may be input to a maximum-likelihood detection and classification algorithm such as is widely used in radar signal processing. The  
15 detection processing uses matched filtering of BCS observed in mass-basecount space and allows for detection and subtraction of signatures from known, harmless organisms, and for detection of unknown bioagent threats. Comparison of newly observed bioagents to known bioagents is also possible, for estimation of threat level, by comparing their BCS to those of  
20 known organisms and to known forms of pathogenicity enhancement, such as insertion of antibiotic resistance genes or toxin genes.

Processing may end with a Bayesian classifier using log likelihood ratios developed from the observed signals and average background levels. The program emphasizes performance predictions culminating in probability-of-detection versus probability-of-false-alarm plots for conditions involving complex backgrounds of naturally occurring organisms  
25 and environmental contaminants. Matched filters consist of a priori expectations of signal values given the set of primers used for each of the bioagents. A genomic sequence database

(e.g. GenBank) is used to define the mass basecount matched filters. The database contains known threat agents and benign background organisms. The latter is used to estimate and subtract the signature produced by the background organisms. A maximum likelihood detection of known background organisms is implemented using matched filters and a running-sum estimate of the noise covariance. Background signal strengths are estimated and used along with the matched filters to form signatures which are then subtracted. the maximum likelihood process is applied to this "cleaned up" data in a similar manner employing matched filters for the organisms and a running-sum estimate of the noise-covariance for the cleaned up data.

In one embodiment, a strategy to "triangulate" each organism by measuring signals from multiple core genes is used to reduce false negative and false positive signals, and enable reconstruction of the origin or hybrid or otherwise engineered bioagents. After identification of multiple core genes, alignments are created from nucleic acid sequence databases. The alignments are then analyzed for regions of conservation and variation, and potential primer binding sites flanking variable regions are identified. Next, amplification target regions for signature analysis are selected which distinguishes organisms based on specific genomic differences (i.e., base composition). For example, detection of signatures for the three part toxin genes typical of *B. anthracis* (Bowen *et al.*, *J. Appl. Microbiol.*, **1999**, 87, 270-278) in the absence of the expected signatures from the *B. anthracis* genome would suggest a genetic engineering event.

Once the molecular mass or base composition signature of the amplification product is determined, it is compared to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known organisms. A match identifies a particular unknown bioagent.

The present invention is also directed to a database having cell-data positional significance. The database comprises at least a first table of a plurality of data-containing cells. The first table is organized into at least a first row and a second row, each row having columns and data-containing cells. The data-containing cells have an alignment with at least one other row for differentiating aligned from non-aligned data-containing cells. The differentiation in alignment of the data-containing cells designates a structural feature of a polymer present in a sample. In some embodiments, the alignment is a vertical alignment according to base pair homology along a linear segment within each polymer. In some embodiments, the vertical alignment further aligns cell-data according to inter-species

conserved regions. In some embodiments, the structural feature is a bulge or a loop and the polymer is an RNA.

The present method can also be used to detect single nucleotide polymorphisms (SNPs), or multiple nucleotide polymorphisms, rapidly and accurately. A SNP is defined as a single base pair site in the genome that is different from one individual to another. The difference can be expressed either as a deletion, an insertion or a substitution, and is frequently linked to a disease state. Because they occur every 100-1000 base pairs, SNPs are the most frequently bound type of genetic marker in the human genome.

For example, sickle cell anemia results from an A-T transition, which encodes a valine rather than a glutamic acid residue. Oligonucleotide primers may be designed such that they bind to sequences that flank a SNP site, followed by nucleotide amplification and mass determination of the amplified product. Because the molecular masses of the resulting product from an individual who does not have sickle cell anemia is different from that of the product from an individual who has the disease, the method can be used to distinguish the two individuals. Thus, the method can be used to detect any known SNP in an individual and thus diagnose or determine increased susceptibility to a disease or condition.

In one embodiment, blood is drawn from an individual and peripheral blood mononuclear cells (PBMC) are isolated and simultaneously tested, preferably in a high-throughput screening method, for one or more SNPs using appropriate primers based on the known sequences which flank the SNP region. The National Center for Biotechnology Information maintains a publicly available database of SNPs on the world wide web of the Internet at, for example, "ncbi.nlm.nih.gov/SNP/."

The method of the present invention can also be used for blood typing. The gene encoding A, B or O blood type can differ by four single nucleotide polymorphisms. If the gene contains the sequence CGTGGTGACCCTT (SEQ ID NO:5), antigen A results. If the gene contains the sequence CGTCGTCACCGCTA (SEQ ID NO:6) antigen B results. If the gene contains the sequence CGTGGT-ACCCCTT (SEQ ID NO:7), blood group O results ("-" indicates a deletion). These sequences can be distinguished by designing a single primer pair which flanks these regions, followed by amplification and mass determination.

The present invention is also directed to a service or methods providing information regarding the safety of a water source. A dimensional master database for storing a molecular mass or base composition signature or both, and an identity, and optionally a geographical detail corresponding to a plurality of known bioagents is provided. The master database

stores the molecular mass or base composition signature or both, the identity, and optionally the geographical detail for a plurality of known bioagents. The master database is interrogated with a tracking request of an unknown or known bioagent obtained from a sample to generate a response. The response is delivered from the master database to the requester. In some embodiments, the molecular mass or base composition signature is of a selected portion of the bioagent, the identity comprises a name for the bioagent, and the optional geographic detail comprises at least one geographic locality for the bioagent. In some embodiments, the interrogation comprises contacting nucleic acid from a bioagent obtained from a sample with either i) at least one pair of oligonucleotide primers which hybridize to sequences of the nucleic acid, wherein the sequences flank a variable nucleic acid sequence of the bioagent, or ii) at least one oligonucleotide primer that hybridizes to a sequence of the nucleic acid, wherein a variable nucleic acid sequence of the bioagent is flanked by the primer and a natural stop region of the nucleic acid, amplifying the variable nucleic acid sequence to produce an amplification product, determining the molecular mass or base composition signature of the amplification product; and comparing the molecular mass or base composition signature of the amplification product to one or more molecular masses or base composition signatures of amplification products obtained from a plurality of known organisms, wherein a match identifies a particular bioagent. In some embodiments, the response is delivered through a network. In some embodiments, the network is a local area network, a wide area network, or the internet.

In other embodiments of the invention, all of the aforementioned methods can be used to determine or confirm the absence of a known or unknown bioagent in a sample by detecting the absence of a particular bioagent in a sample. For example; when the molecular mass or base composition is compared to one or more molecular masses or base compositions obtained from a plurality of known organisms and no match is obtained, the absence of a particular bioagent is determined or confirmed. In these methods of determining the absence of a bioagent, a positive control can be used to confirm the integrity of the system. Such positive controls include, but are not limited to addition of a known bioagent to the sample and confirming the presence of the known bioagent in the sample by obtaining a match from the comparison step. Alternately, residual primer signal can be detected, thus indicating that the system integrity is intact.

Methods of identifying bioagents in environmental and product testing are disclosed in U.S. Serial No. 10/660,996 which is commonly owned and which is incorporated herein by reference in its entirety.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following examples serve only to illustrate the invention and are not intended to limit the same.

In order that the invention disclosed herein may be more efficiently understood, examples are provided below. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting the invention in any manner.

Throughout these examples, molecular cloning reactions, and other standard recombinant DNA techniques, were carried out according to methods described in Maniatis et al., *Molecular Cloning - A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press (1989), using commercially available reagents, except where otherwise noted.

## EXAMPLES

### Example 1: Nucleic Acid Isolation and PCR

In one embodiment, nucleic acid is isolated from the organisms and amplified by PCR using standard methods prior to BCS determination by mass spectrometry. Nucleic acid is isolated, for example, by detergent lysis of bacterial cells, centrifugation and ethanol precipitation. Nucleic acid isolation methods are described in, for example, *Current Protocols in Molecular Biology* (Ausubel et al.) and *Molecular Cloning: A Laboratory Manual* (Sambrook et al.). The nucleic acid is then amplified using standard methodology, such as PCR, with primers which bind to conserved regions of the nucleic acid which contain an intervening variable sequence as described below.

### Example 2: Mass spectrometry

**FTICR Instrumentation:** The FTICR instrument is based on a 7 tesla actively shielded superconducting magnet and modified Bruker Daltonics Apex II 70e ion optics and vacuum chamber. The spectrometer is interfaced to a LEAP PAL autosampler and a custom fluidics control system for high throughput screening applications. Samples are analyzed directly from 96-well or 384-well microtiter plates at a rate of about 1 sample/minute. The Bruker data-acquisition platform is supplemented with a lab-built ancillary NT datastation which controls the autosampler and contains an arbitrary waveform generator capable of

generating complex rf-excite waveforms (frequency sweeps, filtered noise, stored waveform inverse Fourier transform (SWIFT), etc.) for sophisticated tandem MS experiments. For oligonucleotides in the 20-30-mer regime typical performance characteristics include mass resolving power in excess of 100,000 (FWHM), low ppm mass measurement errors, and an  
5 operable  $m/z$  range between 50 and 5000  $m/z$ .

**Modified ESI Source:** In sample-limited analyses, analyte solutions are delivered at 150 nL/minute to a 30 mm i.d. fused-silica ESI emitter mounted on a 3-D micromanipulator. The ESI ion optics consists of a heated metal capillary, an rf-only hexapole, a skimmer cone, and an auxiliary gate electrode. The 6.2 cm rf-only hexapole is comprised of 1 mm diameter  
10 rods and is operated at a voltage of 380 Vpp at a frequency of 5 MHz. A lab-built electro-mechanical shutter can be employed to prevent the electrospray plume from entering the inlet capillary unless triggered to the "open" position via a TTL pulse from the data station. When in the "closed" position, a stable electrospray plume is maintained between the ESI emitter and the face of the shutter. The back face of the shutter arm contains an elastomeric seal that  
15 can be positioned to form a vacuum seal with the inlet capillary. When the seal is removed, a 1 mm gap between the shutter blade and the capillary inlet allows constant pressure in the external ion reservoir regardless of whether the shutter is in the open or closed position. When the shutter is triggered, a "time slice" of ions is allowed to enter the inlet capillary and is subsequently accumulated in the external ion reservoir. The rapid response time of the ion  
20 shutter (< 25 ms) provides reproducible, user defined intervals during which ions can be injected into and accumulated in the external ion reservoir.

**Apparatus for Infrared Multiphoton Dissociation** A 25 watt CW CO<sub>2</sub> laser operating at 10.6  $\mu\text{m}$  has been interfaced to the spectrometer to enable infrared multiphoton dissociation (IRMPD) for oligonucleotide sequencing and other tandem MS applications. An  
25 aluminum optical bench is positioned approximately 1.5 m from the actively shielded superconducting magnet such that the laser beam is aligned with the central axis of the magnet. Using standard IR-compatible mirrors and kinematic mirror mounts, the unfocused 3 mm laser beam is aligned to traverse directly through the 3.5 mm holes in the trapping electrodes of the FTICR trapped ion cell and longitudinally traverse the hexapole region of  
30 the external ion guide finally impinging on the skimmer cone. This scheme allows IRMPD to be conducted in an  $m/z$  selective manner in the trapped ion cell (e.g. following a SWIFT isolation of the species of interest), or in a broadband mode in the high pressure region of the

external ion reservoir where collisions with neutral molecules stabilize IRMPD-generated metastable fragment ions resulting in increased fragment ion yield and sequence coverage.

### Example 3: Identification of Bioagents

5           Table 2 shows a small cross section of a database of calculated molecular masses for over 9 primer sets and approximately 30 organisms. The primer sets were derived from rRNA alignment. Examples of regions from rRNA consensus alignments are shown in Figures 1A-1C. Lines with arrows are examples of regions to which intelligent primer pairs for PCR are designed. The primer pairs are >95% conserved in the bacterial sequence  
10       database (currently over 10,000 organisms). The intervening regions are variable in length and/or composition, thus providing the base composition “signature” (BCS) for each organism. Primer pairs were chosen so the total length of the amplified region is less than about 80-90 nucleotides. The label for each primer pair represents the starting and ending base number of the amplified region on the consensus diagram.

15           Included in the short bacterial database cross-section in Table 2 are many well known pathogens/biowarfare agents (shown in bold/red typeface) such as *Bacillus anthracis* or *Yersinia pestis* as well as some of the bacterial organisms found commonly in the natural environment such as *Streptomyces*. Even closely related organisms can be distinguished from each other by the appropriate choice of primers. For instance, two low G+C organisms,  
20       *Bacillus anthracis* and *Staph aureus*, can be distinguished from each other by using the primer pair defined by 16S\_1337 or 23S\_855 ( $\Delta M$  of 4 Da).

**Table 2: Cross Section Of A Database Of Calculated Molecular Masses<sup>1</sup>**

Primer Regions ---->	16S_971	16S_1100	16S_1337	16S_1294	16S_1228	23S_1021	23S_855	23S_193	23S_115
Bug Name									
Acinetobacter calcoaceticus	55619.1	55004	28446.7	35854.9	51295.4	30299	42654	39557.5	54999
<b>Bacillus anthracis</b>	<b>55005</b>	<b>54388</b>	<b>28448</b>	<b>35238</b>	<b>51296</b>	<b>30295</b>	<b>42651</b>	<b>39560</b>	<b>56850</b>
Bacillus cereus	55622.1	54387.9	28447.6	35854.9	51296.4	30295	42651	39560.5	56850.3
Bordetella bronchiseptica	56857.3	51300.4	28446.7	35857.9	51307.4	30299	42653	39559.5	51920.5
Borrelia burgdorferi	56231.2	55621.1	28440.7	35852.9	51295.4	30297	42029.9	38941.4	52524.6
<b>Brucella abortus</b>	<b>58098</b>	<b>55011</b>	<b>28448</b>	<b>35854</b>	<b>50683</b>				
Campylobacter jejuni	58088.5	54386.9	29061.8	35856.9	50674.3	30294	42032.9	39558.5	45732.5
<b>Chlamydia pneumoniae</b>	<b>55000</b>	<b>55007</b>	<b>29063</b>	<b>35855</b>	<b>50676</b>	<b>30295</b>	<b>42036</b>	<b>38941</b>	<b>56230</b>
<b>Clostridium botulinum</b>	<b>55006</b>	<b>53767</b>	<b>28445</b>	<b>35855</b>	<b>51291</b>	<b>30300</b>	<b>42656</b>	<b>39562</b>	<b>54999</b>
Clostridium difficile	56855.3	54386.9	28444.7	35853.9	51296.4	30294	41417.8	39556.5	55612.2
Enterococcus faecalis	55620.1	54387.9	28447.6	35858.9	51296.4	30297	42652	39559.5	56849.3
<b>Escherichia coli</b>	<b>55622</b>	<b>55009</b>	<b>28445</b>	<b>35857</b>	<b>51301</b>	<b>30301</b>	<b>42656</b>	<b>39562</b>	<b>54999</b>
<b>Francisella tularensis</b>	<b>53769</b>	<b>54385</b>	<b>28445</b>	<b>35856</b>	<b>51298</b>				
Haemophilus influenzae	55620.1	55006	28444.7	35855.9	51298.4	30298	42656	39560.5	55613.1
Klebsiella pneumoniae	55622.1	55008	28442.7	35856.9	51297.4	30300	42655	39562.5	55000
<b>Legionella pneumophila</b>	<b>55618</b>	<b>55626</b>	<b>28446</b>	<b>35857</b>	<b>51303</b>				
Mycobacterium avium	54390.9	55631.1	29064.8	35858.9	51915.5	30298	42656	38942.4	56241.2
Mycobacterium leprae	54389.9	55629.1	29064.8	35860.9	51917.5	30298	42656	39559.5	56240.2
Mycobacterium tuberculosis	54390.9	55629.1	29064.8	35860.9	51301.4	30299	42656	39560.5	56243.2
Mycoplasma genitalium	53143.7	45115.4	29061.8	35854.9	50671.3	30294	43264.1	39558.5	56842.4
Mycoplasma pneumoniae	53143.7	45118.4	29061.8	35854.9	50673.3	30294	43264.1	39559.5	56843.4
Neisseria gonorrhoeae	55627.1	54389.9	28445.7	35855.9	51302.4	30300	42649	39561.5	55000
<b>Pseudomonas aeruginosa</b>	<b>55623</b>	<b>55010</b>	<b>28443</b>	<b>35858</b>	<b>51301</b>	<b>30298</b>	<b>43272</b>	<b>39558</b>	<b>55619</b>
<b>Rickettsia prowazekii</b>	<b>58093</b>	<b>55621</b>	<b>28448</b>	<b>35853</b>	<b>50677</b>	<b>30293</b>	<b>42650</b>	<b>39559</b>	<b>53139</b>
<b>Rickettsia rickettsii</b>	<b>58094</b>	<b>55623</b>	<b>28448</b>	<b>35853</b>	<b>50679</b>	<b>30293</b>	<b>42648</b>	<b>39559</b>	<b>53755</b>
<b>Salmonella typhimurium</b>	<b>55622</b>	<b>55005</b>	<b>28445</b>	<b>35857</b>	<b>51301</b>	<b>30301</b>	<b>42658</b>		
<b>Shigella dysenteriae</b>	<b>55623</b>	<b>55009</b>	<b>28444</b>	<b>35857</b>	<b>51301</b>				
Staphylococcus aureus	56854.3	54386.9	28443.7	35852.9	51294.4	30298	42655	39559.5	57466.4
Streptomyces	54389.9	59341.6	29063.8	35858.9	51300.4			39563.5	56864.3
Trypanosoma pallidum	56245.2	55631.1	28445.7	35851.9	51297.4	30299	42034.9	38939.4	57473.4
<b>Vibrio cholerae</b>	<b>55625</b>	<b>55626</b>	<b>28443</b>	<b>35857</b>	<b>52536</b>	<b>29063</b>	<b>30303</b>	<b>35241</b>	<b>50675</b>
Vibrio parahaemolyticus	54384.9	55626.1	28444.7	34620.7	50064.2				
<b>Yersinia pestis</b>	<b>55620</b>	<b>55626</b>	<b>28443</b>	<b>35857</b>	<b>51299</b>				

<sup>1</sup>Molecular mass distribution of PCR amplified regions for a selection of organisms (rows) across various primer pairs (columns). Pathogens are shown in **bold**. Empty cells indicate presently incomplete or missing data.

Figure 6 shows the use of ESI-FT-ICR MS for measurement of exact mass. The spectra from 46mer PCR products originating at position 1337 of the 16S rRNA from *S. aureus* (upper) and *B. anthracis* (lower) are shown. These data are from the region of the spectrum containing signals from the  $[M-8H]^+^{8-}$  charge states of the respective 5'-3' strands.

The two strands differ by two (AT→CG) substitutions, and have measured masses of 14206.396 and  $14208.373 \pm 0.010$  Da, respectively. The possible base compositions derived from the masses of the forward and reverse strands for the *B. anthracis* products are listed in Table 3.

**Table 3: Possible base composition for *B. anthracis* products**

Calc. Mass	Error	Base Comp.
14208.2935	0.079520	A1 G17 C10 T18
14208.3160	0.056980	A1 G20 C15 T10
14208.3386	0.034440	A1 G23 C20 T2



14208.3074	0.065560	A6 G11 C3 T26
14208.3300	0.043020	A6 G14 C8 T18
14208.3525	0.020480	A6 G17 C13 T10
14208.3751	0.002060	A6 G20 C18 T2
14208.3439	0.029060	A11 G8 C1 T26
14208.3665	0.006520	A11 G11 C6 T18
<b>14208.3890</b>	<b>0.016020</b>	<b>A11 G14 C11 T10</b>
14208.4116	0.038560	A11 G17 C16 T2
14208.4030	0.029980	A16 G8 C4 T18
14208.4255	0.052520	A16 G11 C9 T10
14208.4481	0.075060	A16 G14 C14 T2
14208.4395	0.066480	A21 G5 C2 T18
14208.4620	0.089020	A21 G8 C7 T10
14079.2624	0.080600	A0 G14 C13 T19
14079.2849	0.058060	A0 G17 C18 T11
14079.3075	0.035520	A0 G20 C23 T3
14079.2538	0.089180	A5 G5 C1 T35
14079.2764	0.066640	A5 G8 C6 T27
14079.2989	0.044100	A5 G11 C11 T19
14079.3214	0.021560	A5 G14 C16 T11
14079.3440	0.000980	A5 G17 C21 T3
14079.3129	0.030140	A10 G5 C4 T27
14079.3354	0.007600	A10 G8 C9 T19
<b>14079.3579</b>	<b>0.014940</b>	<b>A10 G11 C14 T11</b>
14079.3805	0.037480	A10 G14 C19 T3
14079.3494	0.006360	A15 G2 C2 T27
14079.3719	0.028900	A15 G5 C7 T19
14079.3944	0.051440	A15 G8 C12 T11
14079.4170	0.073980	A15 G11 C17 T3
14079.4084	0.065400	A20 G2 C5 T19
14079.4309	0.087940	A20 G5 C10 T13

Among the 16 compositions for the forward strand and the 18 compositions for the reverse strand that were calculated, only one pair (shown in **bold**) are complementary, corresponding to the actual base compositions of the *B. anthracis* PCR products.

#### 5 **Example 4: BCS of Region from *Bacillus anthracis* and *Bacillus cereus***

A conserved *Bacillus* region from *B. anthracis* (A<sub>14</sub>G<sub>9</sub>C<sub>14</sub>T<sub>9</sub>) and *B. cereus* (A<sub>15</sub>G<sub>9</sub>C<sub>13</sub>T<sub>9</sub>) having a C to A base change was synthesized and subjected to ESI-TOF MS. The results are shown in Figure 7 in which the two regions are clearly distinguished using the method of the present invention (MW=14072.26 vs. 14096.29).

10

#### **Example 5: Identification of additional bioagents**

In other examples of the present invention, the pathogen *Vibrio cholera* can be distinguished from *Vibrio parahaemolyticus* with  $\Delta M > 600$  Da using one of three 16S primer sets shown in Table 2 (16S\_971, 16S\_1228 or 16S\_1294) as shown in Table 4. The two mycoplasma species in the list (*M. genitalium* and *M. pneumoniae*) can also be distinguished from each other, as can the three mycobacteriae. While the direct mass measurements of amplified products can identify and distinguish a large number of organisms, measurement of the base composition signature provides dramatically enhanced resolving power for closely related organisms. In cases such as *Bacillus anthracis* and *Bacillus cereus* that are virtually indistinguishable from each other based solely on mass differences, compositional analysis or fragmentation patterns are used to resolve the differences. The single base difference between the two organisms yields different fragmentation patterns, and despite the presence of the ambiguous/unidentified base N at position 20 in *B. anthracis*, the two organisms can be identified.

25 Tables 4a-b show examples of primer pairs from Table 1 which distinguish pathogens from background.

**Table 4a**

Organism name	23S_855	16S_1337	23S_1021
<i>Bacillus anthracis</i>	42650.98	28447.65	30294.98
<i>Staphylococcus aureus</i>	42654.97	28443.67	30297.96

Table 4b

Organism name	16S_971	16S_1294	16S_1228
<i>Vibrio cholerae</i>	55625.09	35856.87	52535.59
<i>Vibrio parahaemolyticus</i>	54384.91	34620.67	50064.19

Table 4 shows the expected molecular weight and base composition of region 16S\_1100-1188 in *Mycobacterium avium* and *Streptomyces sp.*

5

Table 5

Region	Organism name	Length	Molecular weight	Base comp.
16S_1100-1188	<i>Mycobacterium avium</i>	82	25624.1728	A <sub>16</sub> G <sub>32</sub> C <sub>18</sub> T <sub>16</sub>
16S_1100-1188	<i>Streptomyces sp.</i>	96	29904.871	A <sub>17</sub> G <sub>38</sub> C <sub>27</sub> T <sub>14</sub>

Table 5 shows base composition (single strand) results for 16S\_1100-1188 primer amplification reactions different species of bacteria. Species which are repeated in the table (e.g., *Clostridium botulinum*) are different strains which have different base compositions in the 16S\_1100-1188 region.

10

Table 6

Organism name	Base comp.	Organism name	Base comp.
<i>Mycobacterium avium</i>	A <sub>16</sub> G <sub>32</sub> C <sub>18</sub> T <sub>16</sub>	<i>Vibrio cholerae</i>	A <sub>23</sub> G <sub>30</sub> C <sub>21</sub> T <sub>16</sub>
<i>Streptomyces sp.</i>	A <sub>17</sub> G <sub>38</sub> C <sub>27</sub> T <sub>14</sub>	<i>Aeromonas hydrophila</i>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Ureaplasma urealyticum</i>	A <sub>18</sub> G <sub>30</sub> C <sub>17</sub> T <sub>17</sub>	<i>Aeromonas salmonicida</i>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Streptomyces sp.</i>	A <sub>19</sub> G <sub>36</sub> C <sub>24</sub> T <sub>18</sub>	<i>Mycoplasma genitalium</i>	A <sub>24</sub> G <sub>19</sub> C <sub>12</sub> T <sub>18</sub>
<i>Mycobacterium leprae</i>	A <sub>20</sub> G <sub>32</sub> C <sub>22</sub> T <sub>16</sub>	<i>Clostridium botulinum</i>	A <sub>24</sub> G <sub>25</sub> C <sub>18</sub> T <sub>20</sub>
<i>M. tuberculosis</i>	A <sub>20</sub> G <sub>33</sub> C <sub>21</sub> T <sub>16</sub>	<i>Bordetella bronchiseptica</i>	A <sub>24</sub> G <sub>26</sub> C <sub>19</sub> T <sub>14</sub>
<i>Nocardia asteroides</i>	A <sub>20</sub> G <sub>33</sub> C <sub>21</sub> T <sub>16</sub>	<i>Francisella tularensis</i>	A <sub>24</sub> G <sub>26</sub> C <sub>19</sub> T <sub>19</sub>
<i>Fusobacterium necroforum</i>	A <sub>21</sub> G <sub>26</sub> C <sub>22</sub> T <sub>18</sub>	<i>Bacillus anthracis</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>18</sub>
<i>Listeria monocytogenes</i>	A <sub>21</sub> G <sub>27</sub> C <sub>19</sub> T <sub>19</sub>	<i>Campylobacter jejuni</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>18</sub>
<i>Clostridium botulinum</i>	A <sub>21</sub> G <sub>27</sub> C <sub>19</sub> T <sub>21</sub>	<i>Staphylococcus aureus</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>18</sub>
<i>Neisseria gonorrhoeae</i>	A <sub>21</sub> G <sub>28</sub> C <sub>21</sub> T <sub>18</sub>	<i>Helicobacter pylori</i>	A <sub>24</sub> G <sub>26</sub> C <sub>20</sub> T <sub>19</sub>

<i>Bartonella quintana</i>	A <sub>21</sub> G <sub>30</sub> C <sub>22</sub> T <sub>16</sub>	<i>Helicobacter pylori</i>	A <sub>24</sub> G <sub>26</sub> C <sub>21</sub> T <sub>18</sub>
<i>Enterococcus faecalis</i>	A <sub>22</sub> G <sub>27</sub> C <sub>20</sub> T <sub>19</sub>	<i>Moraxella catarrhalis</i>	A <sub>24</sub> G <sub>26</sub> C <sub>23</sub> T <sub>16</sub>
<i>Bacillus megaterium</i>	A <sub>22</sub> G <sub>28</sub> C <sub>20</sub> T <sub>18</sub>	<i>Haemophilus influenzae Rd</i>	A <sub>24</sub> G <sub>28</sub> C <sub>20</sub> T <sub>17</sub>
<i>Bacillus subtilis</i>	A <sub>22</sub> G <sub>28</sub> C <sub>21</sub> T <sub>17</sub>	<i>Chlamydia trachomatis</i>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Pseudomonas aeruginosa</i>	A <sub>22</sub> G <sub>29</sub> C <sub>23</sub> T <sub>15</sub>	<i>Chlamydophila pneumoniae</i>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Legionella pneumophila</i>	A <sub>22</sub> G <sub>32</sub> C <sub>20</sub> T <sub>16</sub>	<i>C. pneumonia AR39</i>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Mycoplasma pneumoniae</i>	A <sub>23</sub> G <sub>20</sub> C <sub>14</sub> T <sub>16</sub>	<i>Pseudomonas putida</i>	A <sub>24</sub> G <sub>29</sub> C <sub>21</sub> T <sub>16</sub>
<i>Clostridium botulinum</i>	A <sub>23</sub> G <sub>26</sub> C <sub>20</sub> T <sub>19</sub>	<i>Proteus vulgaris</i>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Enterococcus faecium</i>	A <sub>23</sub> G <sub>26</sub> C <sub>21</sub> T <sub>18</sub>	<i>Yersinia pestis</i>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Acinetobacter calcoaceti</i>	A <sub>23</sub> G <sub>26</sub> C <sub>21</sub> T <sub>19</sub>	<i>Yersinia pseudotuberculosis</i>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Leptospira borgpeterseni</i>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>	<i>Clostridium botulinum</i>	A <sub>25</sub> G <sub>24</sub> C <sub>18</sub> T <sub>21</sub>
<i>Leptospira interrogans</i>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>	<i>Clostridium tetani</i>	A <sub>25</sub> G <sub>25</sub> C <sub>18</sub> T <sub>20</sub>
<i>Clostridium perfringens</i>	A <sub>23</sub> G <sub>27</sub> C <sub>19</sub> T <sub>19</sub>	<i>Francisella tularensis</i>	A <sub>25</sub> G <sub>25</sub> C <sub>19</sub> T <sub>19</sub>
<b><i>Bacillus anthracis</i></b>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>	<i>Acinetobacter calcoacetic</i>	A <sub>25</sub> G <sub>26</sub> C <sub>20</sub> T <sub>19</sub>
<b><i>Bacillus cereus</i></b>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>	<i>Bacteriodes fragilis</i>	A <sub>25</sub> G <sub>27</sub> C <sub>16</sub> T <sub>22</sub>
<b><i>Bacillus thuringiensis</i></b>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>	<i>Chlamydophila psittaci</i>	A <sub>25</sub> G <sub>27</sub> C <sub>21</sub> T <sub>16</sub>
<i>Aeromonas hydrophila</i>	A <sub>23</sub> G <sub>29</sub> C <sub>21</sub> T <sub>16</sub>	<i>Borrelia burgdorferi</i>	A <sub>25</sub> G <sub>29</sub> C <sub>17</sub> T <sub>19</sub>
<i>Escherichia coli</i>	A <sub>23</sub> G <sub>29</sub> C <sub>21</sub> T <sub>16</sub>	<i>Streptobacillus monilifor</i>	A <sub>26</sub> G <sub>26</sub> C <sub>20</sub> T <sub>16</sub>
<i>Pseudomonas putida</i>	A <sub>23</sub> G <sub>29</sub> C <sub>21</sub> T <sub>17</sub>	<i>Rickettsia prowazekii</i>	A <sub>26</sub> G <sub>28</sub> C <sub>18</sub> T <sub>18</sub>
<b><i>Escherichia coli</i></b>	A <sub>23</sub> G <sub>29</sub> C <sub>22</sub> T <sub>15</sub>	<i>Rickettsia rickettsii</i>	A <sub>26</sub> G <sub>28</sub> C <sub>20</sub> T <sub>16</sub>
<b><i>Shigella dysenteriae</i></b>	A <sub>23</sub> G <sub>29</sub> C <sub>22</sub> T <sub>15</sub>	<i>Mycoplasma mycoides</i>	A <sub>28</sub> G <sub>23</sub> C <sub>16</sub> T <sub>20</sub>

The same organism having different base compositions are different strains. Groups of organisms which are highlighted or in italics have the same base compositions in the amplified region. Some of these organisms can be distinguished using multiple primers. For example, *Bacillus anthracis* can be distinguished from *Bacillus cereus* and *Bacillus thuringiensis* using the primer 16S\_971-1062 (Table 6). Other primer pairs which produce unique base composition signatures are shown in Table 6 (bold). Clusters containing very similar threat and ubiquitous non-threat organisms (e.g. *anthracis* cluster) are distinguished at high resolution with focused sets of primer pairs. The known biowarfare agents in Table 6 are *Bacillus anthracis*, *Yersinia pestis*, *Francisella tularensis* and *Rickettsia prowazekii*.

Table 7

Organism	16S_971-1062	16S_1228-1310	16S_1100-1188
<i>Aeromonas hydrophila</i>	A <sub>21</sub> G <sub>29</sub> C <sub>22</sub> T <sub>20</sub>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Aeromonas salmonicida</i>	A <sub>21</sub> G <sub>29</sub> C <sub>22</sub> T <sub>20</sub>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>31</sub> C <sub>21</sub> T <sub>15</sub>
<i>Bacillus anthracis</i>	<b>A<sub>21</sub>G<sub>27</sub>C<sub>22</sub>T<sub>22</sub></b>	A <sub>24</sub> G <sub>22</sub> C <sub>19</sub> T <sub>18</sub>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>
<i>Bacillus cereus</i>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>24</sub> G <sub>22</sub> C <sub>19</sub> T <sub>18</sub>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>
<i>Bacillus thuringiensis</i>	A <sub>22</sub> G <sub>27</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>24</sub> G <sub>22</sub> C <sub>19</sub> T <sub>18</sub>	A <sub>23</sub> G <sub>27</sub> C <sub>20</sub> T <sub>18</sub>
<i>Chlamydia trachomatis</i>	<b>A<sub>22</sub>G<sub>26</sub>C<sub>20</sub>T<sub>23</sub></b>	<b>A<sub>24</sub>G<sub>23</sub>C<sub>19</sub>T<sub>16</sub></b>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Chlamydia pneumoniae</i> AR39	A <sub>26</sub> G <sub>23</sub> C <sub>20</sub> T <sub>22</sub>	A <sub>26</sub> G <sub>22</sub> C <sub>16</sub> T <sub>18</sub>	A <sub>24</sub> G <sub>28</sub> C <sub>21</sub> T <sub>16</sub>
<i>Leptospira borgpetersenii</i>	A <sub>22</sub> G <sub>26</sub> C <sub>20</sub> T <sub>21</sub>	A <sub>22</sub> G <sub>25</sub> C <sub>21</sub> T <sub>15</sub>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>
<i>Leptospira interrogans</i>	A <sub>22</sub> G <sub>26</sub> C <sub>20</sub> T <sub>21</sub>	A <sub>22</sub> G <sub>25</sub> C <sub>21</sub> T <sub>15</sub>	A <sub>23</sub> G <sub>26</sub> C <sub>24</sub> T <sub>15</sub>
<i>Mycoplasma genitalium</i>	A <sub>28</sub> G <sub>23</sub> C <sub>15</sub> T <sub>22</sub>	<b>A<sub>30</sub>G<sub>18</sub>C<sub>15</sub>T<sub>19</sub></b>	<b>A<sub>24</sub>G<sub>19</sub>C<sub>12</sub>T<sub>18</sub></b>
<i>Mycoplasma pneumoniae</i>	A <sub>28</sub> G <sub>23</sub> C <sub>15</sub> T <sub>22</sub>	<b>A<sub>27</sub>G<sub>19</sub>C<sub>16</sub>T<sub>20</sub></b>	<b>A<sub>23</sub>G<sub>20</sub>C<sub>14</sub>T<sub>16</sub></b>
<i>Escherichia coli</i>	<b>A<sub>22</sub>G<sub>28</sub>C<sub>20</sub>T<sub>22</sub></b>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>29</sub> C <sub>22</sub> T <sub>15</sub>
<i>Shigella dysenteriae</i>	<b>A<sub>22</sub>G<sub>28</sub>C<sub>21</sub>T<sub>21</sub></b>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>13</sub>	A <sub>23</sub> G <sub>29</sub> C <sub>22</sub> T <sub>15</sub>
<i>Proteus vulgaris</i>	<b>A<sub>23</sub>G<sub>26</sub>C<sub>22</sub>T<sub>21</sub></b>	<b>A<sub>26</sub>G<sub>24</sub>C<sub>19</sub>T<sub>14</sub></b>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Yersinia pestis</i>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>25</sub> G <sub>24</sub> C <sub>20</sub> T <sub>14</sub>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Yersinia pseudotuberculosis</i>	A <sub>24</sub> G <sub>25</sub> C <sub>21</sub> T <sub>22</sub>	A <sub>25</sub> G <sub>24</sub> C <sub>20</sub> T <sub>14</sub>	A <sub>24</sub> G <sub>30</sub> C <sub>21</sub> T <sub>15</sub>
<i>Francisella tularensis</i>	<b>A<sub>20</sub>G<sub>25</sub>C<sub>21</sub>T<sub>23</sub></b>	<b>A<sub>23</sub>G<sub>26</sub>C<sub>17</sub>T<sub>17</sub></b>	<b>A<sub>24</sub>G<sub>26</sub>C<sub>19</sub>T<sub>19</sub></b>
<i>Rickettsia prowazekii</i>	<b>A<sub>21</sub>G<sub>26</sub>C<sub>24</sub>T<sub>25</sub></b>	<b>A<sub>24</sub>G<sub>23</sub>C<sub>16</sub>T<sub>19</sub></b>	<b>A<sub>26</sub>G<sub>28</sub>C<sub>18</sub>T<sub>18</sub></b>
<i>Rickettsia rickettsii</i>	<b>A<sub>21</sub>G<sub>26</sub>C<sub>25</sub>T<sub>24</sub></b>	<b>A<sub>24</sub>G<sub>24</sub>C<sub>17</sub>T<sub>17</sub></b>	<b>A<sub>26</sub>G<sub>28</sub>C<sub>20</sub>T<sub>16</sub></b>

The sequence of *B. anthracis* and *B. cereus* in region 16S\_971 is shown below. Shown in bold is the single base difference between the two species which can be detected using the methods of the present invention. *B. anthracis* has an ambiguous base at position 20.

*B. anthracis*\_16S\_971

GCGAAGAACCUUACCAGGUMUUGACAUCCUCUGACAACCCUAGAGAUAGGGCU  
UCUCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:1)

*B.cereus*\_16S\_971

GCGAAGAACCUUACCAGGUCUUGACAUCCUCUGAAAACCCUAGAGAUAGGGCU  
UCUCCUUCGGGAGCAGAGUGACAGGUGGUGCAUGGUU (SEQ ID NO:2)

**Example 6: ESI-TOF MS of *sspE* 56-mer Plus Calibrant**

5           The mass measurement accuracy that can be obtained using an internal mass  
standard in the ESI-MS study of PCR products is shown in Fig.8. The mass standard was a  
20-mer phosphorothioate oligonucleotide added to a solution containing a 56-mer PCR  
product from the *B. anthracis* spore coat protein *sspE*. The mass of the expected PCR product  
distinguishes *B. anthracis* from other species of *Bacillus* such as *B. thuringiensis* and *B.*  
10 *cereus*.

**Example 7: *B. anthracis* ESI-TOF Synthetic 16S\_1228 Duplex**

          An ESI-TOF MS spectrum was obtained from an aqueous solution containing 5  $\mu$ M  
each of synthetic analogs of the expected forward and reverse PCR products from the  
15 nucleotide 1228 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 9) show that the  
molecular weights of the forward and reverse strands can be accurately determined and easily  
distinguish the two strands. The  $[M-21H^+]^{21-}$  and  $[M-20H^+]^{20-}$  charge states are shown.

**Example 8: ESI-FTICR-MS of Synthetic *B. anthracis* 16S\_1337 46 Base Pair Duplex**

20           An ESI-FTICR-MS spectrum was obtained from an aqueous solution containing 5  
 $\mu$ M each of synthetic analogs of the expected forward and reverse PCR products from the  
nucleotide 1337 region of the *B. anthracis* 16S rRNA gene. The results (Fig. 10) show that  
the molecular weights of the strands can be distinguished by this method. The  $[M-16H^+]^{16-}$   
through  $[M-10H^+]^{10-}$  charge states are shown. The insert highlights the resolution that can be  
25 realized on the FTICR-MS instrument, which allows the charge state of the ion to be  
determined from the mass difference between peaks differing by a single  $^{13}C$  substitution.

**Example 9: ESI-TOF MS of 56-mer Oligonucleotide from *saspB* Gene of *B. anthracis*  
with Internal Mass Standard**

30           ESI-TOF MS spectra were obtained on a synthetic 56-mer oligonucleotide (5  $\mu$ M)  
from the *saspB* gene of *B. anthracis* containing an internal mass standard at an ESI of 1.7  
 $\mu$ L/min as a function of sample consumption. The results (Fig. 11) show that the signal to

noise is improved as more scans are summed, and that the standard and the product are visible after only 100 scans.

**Example 10: ESI-TOF MS of an Internal Standard with Tributylammonium (TBA)-trifluoroacetate (TFA) Buffer**

An ESI-TOF-MS spectrum of a 20-mer phosphorothioate mass standard was obtained following addition of 5 mM TBA-TFA buffer to the solution. This buffer strips charge from the oligonucleotide and shifts the most abundant charge state from  $[M-8H^+]^{8-}$  to  $[M-3H^+]^{3-}$  (Fig. 12).

**Example 11: Master Database Comparison**

The molecular masses obtained through Examples 1-10 are compared to molecular masses of known bioagents stored in a master database to obtain a high probability matching molecular mass.

**Example 12: Master Data Base Interrogation over the Internet**

The same procedure as in Example 11 is followed except that the local computer did not store the Master database. The Master database is interrogated over an internet connection, searching for a molecular mass match.

**Example 13: Master Database Updating**

The same procedure as in example 11 is followed except the local computer is connected to the internet and has the ability to store a master database locally. The local computer system periodically, or at the user's discretion, interrogates the Master database, synchronizing the local master database with the global Master database. This provides the current molecular mass information to both the local database as well as to the global Master database. This further provides more of a globalized knowledge base.

**Example 14: Global Database Updating**

The same procedure as in example 13 is followed except there are numerous such local stations throughout the world. The synchronization of each database adds to the diversity of information and diversity of the molecular masses of known bioagents.

Various modifications of the invention, in addition to those described herein, will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. Each reference cited in the present application is incorporated herein by reference in its entirety.